Modeling Qualitative Judgments in Bayesian Networks

José Luis Galán Caballero

Submitted for the degree of Doctor of Philosophy

Queen Mary, University of London

2008

Abstract

Although Bayesian Networks (BNs) are increasingly being used to solve real world problems [47], their use is still constrained by the difficulty of constructing the node probability tables (NPTs). A key challenge is to construct relevant NPTs using the minimal amount of expert elicitation, recognising that it is rarely cost-effective to elicit *complete* sets of probability values.

This thesis describes an approach to defining NPTs for a large class of commonly occurring nodes called *ranked nodes*. This approach is based on the doubly truncated Normal distribution with a central tendency that is invariably a type of a weighted function of the parent nodes.

We demonstrate through two examples how to build large probability tables using the *ranked nodes* approach. Using this approach we are able to build the large probability tables needed to capture the complex models coming from assessing firm's risks in the safety or finance sector.

The aim of the first example with the National Air-Traffic Services(NATS) is to show that using this approach we can model the impact of the organisational factors in avoiding mid-air aircraft collisions. The resulting model was validated by NATS and helped managers to assess the efficiency of the company handling risks and thus, control the likelihood of air-traffic incidents. In the second example, we use BN models to capture the operational risk (OpRisk) in financial institutions. The novelty of this approach is the use of causal reasoning as a means to reduce the uncertainty surrounding this type of risk. This model was validated against the Basel framework [160], which is the emerging international standard regulation governing how financial institutions assess OpRisks.

Contents

To my brother Manuel

Acknowledgements

I am very thankful for the help of my tutors: Prof. Martin Neil and Prof. Norman Fenton whose invaluable guidance and support made possible this dissertation. I am also indebted to my colleagues Dr. Tassos Tombros, Dr. Fabrizzio Esmeraldi, Dr. Pablo Amelin, Dr. Christine Eckes and Terry Conway for having reviewed and commented this thesis as well as to my esteemed colleague Adrian Joseph for his wise advice.

This thesis was supported in part by the EPSRC funded SCORE project (Sensing Changes in Operational Risk Exposure), which was set up in 2001 to investigate how the risk posed by poor organisational culture might be measured and monitored.

Chapter 1

Preface

1.1 Introduction

Since the 90's, the complexity of the BNs has increased noticeably [47]. Mahoney and Laskey [198] commented, early in that decade that a network with over a 100 nodes (e.g. Pathfinder with 109 [51, 96]) was considered a large network. It is only in recent years that BNs are able to handle complex real world networks with thousands of nodes [6, 139, 187, 188]. The developing of large-scale networks have been possible thanks to the modularity of BNs [108], the improvement in the propagation algorithms [108, 112] and to the development of ideas such as Object Oriented Bayesian Networks (OOBNs) [19,122,158]. Also tools such as Hugin [13] and AgenaRisk [134] helps make these models possible.

A consequence of large networks is the need to fill the large probability tables that the complexity of the node's relationship produce [57,61]. There are different ways to deal with these large tables that range from using statistical functions based on data [146], reducing the size of the probability table by making the nodes' state boolean, to interpret the network as a *qualitative network* whose nodes' relationships influence are noted using arithmetic signs, e.g. {+,-} [107, 223] or using techniques such as *divorcing* [112] or *OR-gates/AND-gates* [178].

However, there is no available answer to those networks with large probability tables where the data is scarce and has to be complemented by experts' qualitative knowledge.

The key challenge is to construct relevant NPTs using the minimal amount of expert elicitation, recognising that it is rarely cost-effective to elicit *complete* sets of probability values.

This thesis puts forward an approach that uses experts' qualitative information together with statistical functions to build large probability tables. The methodology defines NPTs for a large class of commonly occurring nodes called *ranked nodes*. This approach is based on the doubly truncated Normal distribution with a central tendency that is invariably a type of weighted function of the parent nodes. This approach has been tested during our research project with NATS and in our OpRisk model and has been formally incorporated into the AgenaRisk tool [134].

1.2 Research Hypothesis

The aim of this thesis is to produce a method to help building large NPTs based largely on expert's knowledge with the following characteristics:

- easy to elicit, in the sense that experts need only to provide a few cues in order to build a NPT;
- accessible method to compute NPT, given that the BN models are used and maintained, in most cases, by people without much knowledge on statistics;
- fast feedback by showing to the expert the results of his assessments on the model through a BN programming tool;
- ablility to develop large BN models requiring less expert time thus producing less costly networks while maintaining the quality of the outcome.

This thesis contributes to a number of fields:

1. to the field of Computer Science in that it provides a new approach to build NPTs from expert elicitation;
2. to the field of knowledge elicitation in that it helps the extraction of experts' judgment in the context of building large BNs;
3. to the field of risk analysis in that it shows how this approach can be applied to assess the risks on the safety critical industries and financial institutions.

1.3 Synopsis

Part I. This part introduces the domains of critical safety risk and operational risks in financial institutions and explains the need for techniques to measure the risk in these domains. These domains are characterised by the lack of data and the availability of expert knowledge. BN is introduced as an alternative to model these risks. We explain Bayes theorem and how it handles uncertainty. This explanation includes the use of graph theory, the concepts of variables' (in)dependencies and the semantic interpretation of BN as causal networks.

Chapter 2. This chapter argues that accidents in the safety critical industries and/or monetary losses in the financial sector derive from the accumulation of minor events that go unnoticed. These minor mishaps form a causal chain that ends in a bigger breakdown: safety breaches resulting in putting people's lives at risk or monetary losses that can end in firm's bankruptcy. Thus the need to reduce or avoid these risks. To this end, we advocate the use of BNs whose characteristics we argue are suitable for this job.

Chapter 3. Introduces Bayes Theorem. This chapter assumes probability theory as the optimal paradigm to handle uncertainty. We explain BN and its interpretation by AI to model uncertainty; this explanation includes Naïve Bayes (NB) and BN. We comment on the contribution of graph theory to develop BNs including the concepts of (in)dependence and directional separation. We also have included a section on Causality. There we comment on the advantages of interpreting BNs as a causal probabilistic net.

Part II. This part highlights the need to rely on domain experts' judgments to inform NPTs in the absence of hard data. It focuses on modeling BNs using knowledge base. It discusses the problems with expert elicitation and, within this context, reviews the techniques currently available to elicit and produce NPTs. This review shows the need for methods to build NPTs based on qualitative information. For that reason we put forward a new approach that, using the minimal amount of expert elicitation, can build large NPTs.

Chapter 4. This chapter focuses on knowledge based Bayesian models. We discuss the subjective nature of these models and the problem of human bias during the elicitation process. We also comment on Ontologies as a possible unifying methodology to build them. We then review the available methods to elicit probability estimations. This review covers methods that use graphical clues, e.g. slide bar, probability wheel; verbal/numerical method (recognising its shortcomings and considering its benefits); statistical techniques that elicit interval's quantiles or distribution parameters.

Chapter 5. In this chapter, we describe a simple approach to defining NPTs for a large class of commonly occurring nodes that we shall call *ranked nodes*. The approach is based on the doubly truncated Normal distribution with a central tendency that is invariably a type of weighted function of the parent nodes. With this approach, we are able to complete a NPT by eliciting a few probability estimations, which in the context of building large probability tables represents a reduction in time and costs.

This approach was put into practice, tested and validated during the development of the National Air-Traffic Service (NATS) model, documented in chapter 6. Using ranked nodes we were able to develop BNs models within time and budget constraints whilst obtaining results that were deemed successful by the experts at NATS.

Part III. This part provides two examples of modeling BNs in the field of safety and finance industries. The first example explains the use of the *ranked node* approach to measure the impact of the factors that intervene on the avoidance of mid-air aircraft collisions. The second example discusses the use of BNs to model financial risk, in particular Operational Risk. It reviews current techniques used to measure this type of risk and puts forward the use of BNs models as an alternative to them.

Chapter 6. The motivation for the BN model in this chapter is to reveal the contributions and the relative importance of "up-stream" factors in air traffic management to the risk of air incidents. The aim of the chapter is to demonstrate that subjective factors such as a company's culture can be measured, hence deriving in better control of safety risks.

This model was developed with the help of the experts at the NATS. With their help we built the topology, probability tables and validated the model's predictions. The experts at NATS described the BNs model as a useful decision making tool for identifying potential safety risks and allocating resources.

Chapter 7. This chapter explains the use of BN models to assess operational risk in financial institutions. The need to identify and measure this risk is given by the introduction of new regulatory measures by the Basel Committee to be effective by 2007. We discuss the content of such regulation and the framework the Committee introduce to assess this type of risk. The aim of this chapter is to compare the current methods against BN models and to show that BNs can offer a viable alternative to them.

To this end, we have built a BN model that combines data and experts' judgment in compliance with the Basel Committee requirements. By combining these information sources, we improve the accuracy of the predictions while increasing our knowledge of the domain area. Furthermore, the BN model explicitly highlights an organisation's weaknesses showing how potential risk emerge from their interaction, thus making the model an essential tool for risk management and for the regulatory body.

Chapter 8. In this chapter we give a summary and we draw the conclusion of this thesis. We comment on the lessons learnt during the development of the NATS project.

Appendix A As part of this thesis we develop a BN tool to help elicit and build the NPTs needed for the NATS project. This tool implements the ideas originated from this project. Ideas such as the ranked node approach. Thanks to this tool we were able to built the NPTs and to run sensitivity analysis. As part of the lessons learnt and the research study we implemented also a Bezier curve, Histograms and Normal distribution to elicit experts' opinions. These tools are now part of the AgenaRisk tool [134].

Appendix B The Safety Attitude Questionnaire (SAQ). This appendix shows the set of questions on SAQ.

Appendix C A **contribution** section has been added to signal my personal contributions to this thesis.

Part I

Problem domain and BNs

Chapter 2

Risks in the Safety industry and Operational Risk in Financial institutions.

2.1 Introduction

Since the 90's we have been more aware of the role that socio-technical factors play within the domain of the safety industry and financial institutions. The Bhopal chemical plant disaster [145], the Ladbroke Grove rail accident [44], the collapse of Barings bank [67], or more recently, the frauds at Allied Irish Bank [10], and ENRON [194] to name but a few, add perspective into this role.

These "breakdowns", whether financial or safety critical, are a burden to society however differently they are perceived. The investigations on these breakdowns often concluded that mishaps are not solely the result of human fallibility, but are supported by organisational features that fail to defend themselves against all-too-human mistakes, slips and (in the case of fraud) malicious acts. From this we can conclude that risk prediction is inextricably entwined with good management practice. That is why the search for improved internal control measures, whether it is in safety hazardous or financial industries, has focused on the underlying organisational structure where those risks occur.

Furthermore, the development of more highly automated technology, the growth of e-commerce, large-scale mergers, the planning of new airports or the re-opening of nuclear plants, all suggest that the risk exposures may be substantial and growing [164].

However, the evaluation of these events cannot readily be handled by traditional statistical methods given that these methods rely on historical data. The novelty of some of the risks together with the lack of data that is associated to safety and operational risk events prevents the use of such methods. Hence traditional statistical techniques often find insufficient data from which to build a sensible model. Moreover, these techniques have failed to capture the interactions between the probability of a risk event and the organisational culture where those risks flourish.

Thus, the need to account for the importance of the management practices in shaping the company's risk profiles, the subjective nature of its measuring and the lack of data regarding these domains, makes the use of Bayesian Networks(BNs) a suitable alternative. Even more so if we consider the importance of risk management in a firm, on one side averting and controlling the threats and on the other transforming threats into a strategic business advantage.

2.2 Background

It is commonly agreed by the academics Reason [190], Whittingham [227], Perrow [182] and Dempsey [53], acknowledged in the field of finance by its regulatory body [162, 163] and by government commissioned reports Turnbull [214], Cadbury [25] that these breakdowns should be attributed to a number of factors among which we find the underlying structure of the organisation. In most cases, we find common factors such as "lack of management commitment and involvement to implement and supervise control risk policies and procedures", "lack of communication between management and employee" or "poor reporting practices", among others, that can be causally linked to the actual failure [213]; they highlight that these human errors may have to do more with the environment where they occur than with the person(s) that caused the mishap[1] [23, 102].

The challenge is to find out the causal sequence of events that lead to the breakdown. In order to better understand this causal chain, we can group the risk an organisation can face into two categories [192]:

- High frequency and low severity events. These are the kind of losses that are usual in the day-to-day running of an organisation, including *direct failures* such as an

[1]See [78] for a literature review on the subject

operative failing to enter the correct day for a transaction and *latent conditions* due to over-complicated procedures or bad designs, causing errors in the long run. To an extent the firm expect to run into these type of risks and this is one of the reasons why these events are not even recorded. For many companies, these are seen as the running costs of the business.

- Low frequency and high impact events. These are unexpected events that occur rarely and with a high level of severity for the firm. These risks may cause the firm to go out of business and/or to put at peril the life of people.

There are different ways to model these risks depending on whether we interpret these two categories as independent or as causally related:

- Both risks are independent and there is no causal link between them. That is, unexpected failures are inherently random events. Which implies that management cannot do much about preventing these risks other than taking insurance.
- Both risks are causally related. That is, unexpected failures are caused by a number of expected ones that went unnoticed over a period of time. Which implies that studying the causal chain of those events can prevent or at least reduce unexpected breakdowns.

In this thesis we agree with J.T. Reason [190, 191] view who sustains that accidents are causally related. He uses the "Swiss-cheese" analogy to illustrate his idea, see Figure 2.1. According to Reason, an accident or *active failure* is the "consequence" of the addition of *latent* failures that were not identified on time. Reason describes four layers of human failure [190]:

- Organizational Influences;
- Unsafe Supervision;
- Preconditions for unsafe acts and
- Unsafe acts.

Shappell and Wiegmann [205] give a more *pragmatical* interpretation to Reason's model. They elaborated the Human Factors Analysis and Classification System (HFACS)

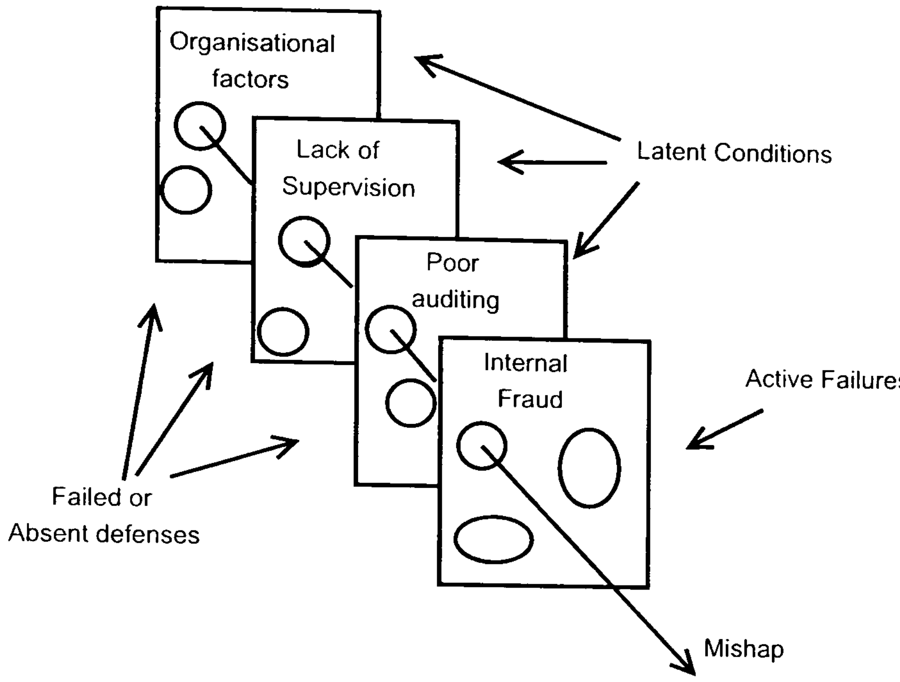


Figure 2.1: This figure shows an accident trajectory passing through corresponding holes in the layers of defense, barriers and safeguards. Adapted from [190].

based on Reason's model of *active* and *latent* failures. They divided these layers into more detail that could be mapped with current operation tasks. For instance, the "unsafe acts" category is divided into two subcategories: Errors (skill based errors, decision errors and perceptual errors) and Violations (routine and exceptional), "preconditions for unsafe acts" or *performance shaping factors* (PSF) are divided into adverse mental state, resource mismanagement, environmental conditions and distractions, among others. It is interesting to note on this point, D O'Hare [157] research's findings. Apart from showing a high correlation between PSF and the likelihood of an aviation mishap, he also comments on the difference between the PSFs found on an accident (e.g. mid-air aircraft collision) with the PSFs found on an incident (e.g. loss of separation between two planes), thus showing the relevance of PSF on an accident or incident. The "Unsafe Supervision" layer is where PSFs take place and this level of supervision is the result of the "Organisational Influences". Using this hierarchical sequence of events motivates management to intervene on the weak points on the chain.

HFACS has, since, been used as a framework to investigate breakdowns [56, 205] on different domains where human error has an input in an accident. For instance, the US Navy/Marine Corps used the same framework to investigate 181 helicopter accidents between 1991 and 1997 35% of those accidents were associated to at least one violation of rules and regulations [205]. HFACS was also used to study 523 aviation accidents in the republic of China Air Force through 1978 to 2002. The results show a key relationship between accidents and errors at the operational and organisational levels that could explain these accidents [131]. Thanden et al [211] obtain similar results after reviewing commercial aviation accidents in US from 1990 to 2000; from the 1322 accidents 781 were identified as a result of human factors after using HFACS; 60 of them were associated with 70 organisational factors. For more examples see [28, 120].

S. A. Shappell and D. A. Wiegmann [205] conclude

> ...HFACS framework has resulted in the improved quality and quantity of information gathered during aviation accident investigation ... In addition, it has proved to be an effective instrument to monitoring the success or failures of specific intervention programmes ...

Hence, we can conclude that finding out the "causal links" between expected and unexpected failures will help managers to identify possible weak points in the system [56, 58, 64, 104, 174]. However, this causal reasoning is not often used to identify and measure rare events because it has proved difficult to find this "link" and to assign a value to it [119]. We argue that this is because traditional statistical methods used to measure "unexpected" risk fails to capture the complexities of the domain of safety risk and operational risk and not because risk is *inherently random* in these domains. Analyzing this causal chain of events would help to separate the events into either "natural" (i.e. random) or "man-made"(i.e. causally explained).

Furthermore, from the point of view of risk management, interpreting unexpected events as single, random events brings a number of problems as the Health and Safety Executive (HSE) comments [95]:

- It often underestimates the true impact of a problem overall. It only focuses on one unexpected risk and ignores the others.
- Treating unexpected risks as being inherently random it undermines the adoption of a precautionary approach based on anticipating and averting losses.
- Measuring only the unexpected losses is inadequate since it often reduces the characteristics of what is a complex issue to a "single number" without analysing other factors such as the organisational factors that contribute to those losses.

In the next section, we review the characteristics of these domains. Studying their features will help us understand why classical statistical methods are not the most suitable candidates to deal with risks in these domains.

2.3 Problem domain: features

The domains of safety and operational risk domains are characterised by the following features:

1. Emergent nature of risks. That is, it is in conjunction with others risk when an incident happens. For example, as we explain in chapter 6, we see that the risk of a mid-air traffic collision depends on a number of factors, e.g. complexity of air space (whether this space is shared, for instance, with the military air force), hardware support (e.g. the Short Term Collision Advice), pilot's skills.

2. Lack of qualitative reporting. There is little information on the causal chain of events leading to a breakdown given the lack of qualitative reporting. As Haas and Kaiser [89] point out

   > ... the major obstacle to a sound data collection process is an underdeveloped risk culture, which for many years has encouraged bank employees to provide overly positive risk reports, and to hide errors and potential losses from their respective superiors ...

3. Sparse data. There is not enough data to build a database from which to extract a pattern of behaviour. Therefore, we need to complement this lack of data with expert's knowledge. Also, the novelty of some risks, e.g. large-scale mergers, new airports, makes way to new possible threats that have not been recorded yet.

4. Subjective information. A good deal of the information available is subjective, e.g. a company's internal control standards, self-assessment questionnaires, key risk indicators, etc.

5. Regulatory body demands. In the case of finance, only recently and due to the demands of the regulatory authorities banks started to document these events with a view to building a database that different financial institutions can share. Hence,

the amount of loss data, required by traditional statistical methods, is not going to be enough, at least not until a few years have passed. And even then, as Fenton [70] comments

> As industry improves there will be less data from which to draw credible statistical estimates.

It is for these reasons we advocate the use of BNs.

2.4 Why Bayesian Networks?

The use of BN to model the risks in the safety and finance industry offers several advantages:

1. Given the emergent nature of the domain and hence its complexity, we could use BN's modularity to breakdown the problem into smaller sub-process/potential factors that could contribute to the overall risk.

2. BNs allow us to integrate qualitative information, expressed as experts' opinions, and quantitative information, however small. As Giudici and Bilotta [84] point out, the combination of both sources of information has an advantage:

   > ... historical measure is statistically precise but backward-looking, as it takes into account past data; the self-assessment measure is statistically imprecise but forward looking, as it may incorporate knowledge on the future strategy of the bank.

3. Using our *ranked node* approach, introduced in chapter 5, we can quantify this qualitative information by asking the experts to score their belief about an event, in a similar fashion as the Scorecards approach already used in many management systems [50].

4. Uncertainty surrounding the domain: Bayes theorem allows us to reason under conditions of uncertainty. The uncertainty produced by the lack of data can be reduced using the experience and knowledge of the experts in the problem domain. Using Bayes Theorem, experts' judgments are expressed in probabilistic terms which could be conditioned on the available data.

5. Uncertainty of the model derived from the difficulty to find all possible factors that contribute on a breakdown. Using BN helps us study the possible models by *maximizing* the most probable explanation.

6. It can perform "what if" scenarios to test models' predictions[2].

7. Graphical representation shows explicitly the operational process involved in a business unit and its expected outcome. In the field of finance this is of particular interest. The methods, required by the financial regulators [164], must be able to perform such an analysis for auditory reasons.

Also from the point of view of risk management and decision making [123]:

- Subjective estimation encourages ownership of the risk.

- De-centralises responsibility: within the organisation we find different subcultures that reflect the particular environment of that part of the company. It is better for them to identify and measure their own risks rather than having one measure to "fit all". For these reason we modeled the impact of organisational culture, in the NATS project, as the result of an organisation-division culture which is how a particular set of people interpret that mainstream culture.

- Better controls: it focuses in control weaknesses thus optimising the risk management.

However, the use of BN leaves us with a number of challenges:

1. Eliciting networks. Given that BN models are a semantic representation of the domain, we face the problem of defining the meaning of qualitative inputs and estimating the value of the quantitative ones.

2. Building large NPTs. One of the biggest obstacles to building large BN models is to provide the NPTs. In some cases, these NPTs need thousands of probability estimations. To elicit each value would at best represent a time consuming and an error prone task and at worst is impossible.

We study these challenges in the following chapters.

[2]This technique is part of the Basel requirements to build risk models [162]

Chapter 3

Bayes Networks

3.1 Introduction

Up until the 1970's, the aim of artificial intelligence (AI) was to produce expert systems (ES) that depended more on reasoning than on domain knowledge. An example of these systems is the game of chess, where the domain knowledge is replaced by a brute force search of possible move combinations.

Although the use of search algorithms proved useful, it was soon realised that ES could not rely entirely on reasoning alone; it also needed to take advantage of the expert's knowledge and experience. As research showed, reasoning plays a relatively small role on the solution of a given problem [81]. For this motive, AI started during the 70's to acknowledge experts' opinion and heuristic knowledge of the domain, and began to incorporate them into the ES.

In the early 1970s, ES started to incorporate expert's knowledge using Classical Logic reasoning. However, it soon became clear that this approach was impractical to handle the uncertainty. In Classical Logic, a proposition is true or false when the "facts" that support it are either true or false respectively. This approach proved to be too coarse to represent real world problems.

A more flexible approach was provided by the interpretation AI made of probability theory based on Bayes theorem to build ES. Other approaches were also used to build ES such as Certainty Factors [55,81,206] based on logic rules, Fuzzy Logic methods [55] based on multi value-logic rules or Dempster-Shafer's theory [81, 204, 209] which incorporates the concept of ignorance to probability reasoning.

In this thesis we are focusing on the use of Bayes theorem to model uncertainty. We view probability theory as the optimal approach to represent uncertainty.

In section 3.2 we explain Bayes theorem and its role developing ES. Section 3.3 introduces graph theory. We included this section as we understand that this is fundamental to understanding the building, elicitation and interpretation of BNs models and as such integral part of this thesis. For the same reason we included in section 3.4 the concepts of (in)dependence and conditional dependency and their role on building BN models. We also explain how to find these (in)dependencies through directional separation. We finish this section comparing two models based on Bayes theorem: using Naïve Bayes (NB) and BN. The aim is to observe the different types of models they both produce and their associated joint probability. In section 3.5, we define BN and the characteristics it has. In section 3.6, we explain the convenience of interpreting BN as a causal network and in section 3.7 we give a summary.

3.2 Bayes Theorem and Reasoning

3.2.1 Bayes Reasoning

The terms objective and subjective probability describe two schools of thought which have been dealing with the interpretation of probability.

The frequentist or "objective" school assigns probabilities to random events according to their frequencies of occurrence. This school assumes that the frequency count of the event reveals the *physical* properties of that event. So, for them it is meaningless to use probabilities to *single events*, which means that only repeatable events have probabilities.

Using Bayesian reasoning, however, we do not need to have a 100 mid-air aircraft collisions to build a frequency count on which to base the probability of mid-air accidents. We know that accidents occur but we are uncertain about their likelihood, for that reason we condition this fact to the evidence we have. We can do this because for Bayesians, randomness refers to the uncertainty surrounding the true value of the parameter and not with its physical random nature. This is possible because probabilities are identified with *degrees of belief* (as long as it follows Cox's rules[1]). The uncertainty about a proposition depends on the degree of belief we assign to it.

As we can see, the term uncertainty can be interpreted as an event subject to random variability or as an event whose knowledge about it is imperfect/incomplete. The first interpretation of uncertainty is the one accepted by the objective school while Bayesians accept both interpretations [36].

The following example illustrates this point: Laplace used Bayesian interpretation of probability theory to estimate the mass of Saturn, given the data from various astronomical observations and relevant background information available, e.g. law of classical mechanics. Another 150 years' accumulation of data has changed Laplace estimate by only 0.63%. According to the frequencist school the laws of probability cannot be applyed to this problem. This is because the mass of Saturn is a constant and not a random variable, therefore, it has no frequency distribution and so the laws of probability cannot be used [208]. However, according to Bayes theory, randomness refers to the uncertainty on the variable's value.

3.2.2 Bayes Theorem

One of the main applications of the Bayes Theorem is to obtain the certitude, in probabilistic terms, of an expert's belief (i.e. hypothesis) h given the supporting evidence e. In probability notation this conditional relationship is written as: $P(h|e)$. Bayes theorem obtains $P(h|e)$ using the *product rule* of probability.

$$P(h|e) = P(h)\frac{P(e|h)}{P(e)} \tag{3.1}$$

This is done because in many situations where we want to compute $P(h|e)$ it turns

[1]Cox gave a set of rules to which degrees of belief can be mapped to probabilities if they are to satisfy logical consistency. Cox axioms are: Taken from [140]

1. Degrees of belief must be ordered; if $B(x) > B(y)$ and $B(y) > B(z)$ then $B(x) > B(z)$. Thus, belief can be mapped onto real numbers.
2. The degree of belief in a proposition x and its negation $\neg x$ are related given a function, $f(x)$ such that $B(x) = f[B(\neg x)]$.
3. The degree of belief in a conjunction of proposition x and y is related to the degree of belief in the conditional proposition $x|y$ and the degree of belief in the proposition y.

In this context, Cox says that probabilities can be used to make assumptions, and to describe inferences given those assumptions.

out that it is difficult to do so directly, yet we might have direct information about $P(e|h)$ [69]. Bayes theorem says that $P(h|e)$, called the *posterior* probability, is related to our *prior* assessment of the hypothesis $P(h)$ and to the *likelihood* of this hypothesis, $P(e|h)$. That is, $P(h|e)$ increases or decreases with $P(h)$ and with $P(e|h)$ [146]. The denominator $P(e)$ normalises the posterior distribution given that all probabilities are subject to the probability of its occurrence in the first place. Thus, given that they all share this common factor we can drop the common denominator from 3.1 and interpret the relationship in terms of proportions; the probability $P(h|e)$ is proportional of the hypothesis h occurring in the first place times the "portion" of this hypothesis occurring among all other hypotheses supported by the same evidence[2]. Using $\propto$ as a proportional symbol we have

$$P(h|e) \propto P(h)P(e|h) \tag{3.2}$$

This relationship becomes clearer if we are indifferent to any particular hypotheses, e.g. choosing a uniform distribution for priors.

$$h_{ML} \equiv argmax P(e|h) \tag{3.3}$$

where *argmax* stands for the probability that maximises, better explains the hypothesis. In this case, we evaluated a hypothesis h in terms of the *Maximum Likelihood* (ML). In equation 3.3, we see more clearly how Bayes theorem conditions the factual evidence on expert's judgment.

Thus, using Bayesian theorem, we can describe by means of conditioning the expert's uncertainty on the available evidence all uncertainties present in the problem [20]. For instance, to diagnose an illness given some observed m symptoms s and n diagnosis d we have equation:

$$P(d_1, \ldots\ldots, d_n|s_1, \ldots., s_m) = \frac{P(s_1, \ldots, s_m|d_1, \ldots., d_n)P(d_1, \ldots., d_n)}{\sum_{d_1 \ldots d_n} P(s_1, \ldots, s_m|d_1, \ldots, d_n)P(d_1, \ldots, d_n)} \tag{3.4}$$

However, this expression is not practical given the number of parameters it re-

[2]Being the evidence the universe of discourse and the hypothesis partitions of that universe which are supported by that evidence. Note that the hypothesis are exclusive and exhaustive.

quires. First, we need all prior values for d and all the conditional probabilities for $(s_1, ..., s_m | d_1, ..., d_n)$. In the case of binary variables, we have 2^n prior probabilities and $2^{m+n} - 1$ conditional probabilities. For example, for a small probabilistic model of 3 diagnosis and 10 symptoms we need around 8.000 parameters. The need for parameters would increase exponentially as the number of symptoms and diagnosis entered increases [55].

So, to reduce the number of parameters needed, it is assumed that the symptoms are *exclusive* and *exhaustive*, that is, only one symptoms is possible at a given time , i.e. not two symptoms can overlap, and all the possible symptoms are present in the model. It is also assumed the symptoms are conditionally independent for a given diagnosis.

$$P(d|s_1, ..., s_m) = \frac{P(s_1|d).....P(s_m|d)P(d)}{\sum_n P(s_1|d_n).....P(s_m|d_n)P(d_n)} \tag{3.5}$$

With these assumptions, we need n prior probabilities *p(d)* and 2m*n conditional probabilities *p(s|d)*, hence, we need, n-1+m*n parameters. So, for 3 diagnosis and 10 symptoms we would need only 32 parameters [55]. See figure 3.1.

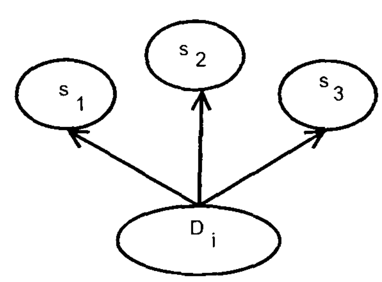


Figure 3.1: Naïve Bayes for three symptoms.

This type of probabilistic model is called Naïve Bayes (NB). In real world applications, these assumptions of independence are hardly tenable. That is the reason why this approach was termed "naïve". NB is still used although circumscribed to the field of medicine for "diagnosis-symptoms" models [62] or as a data classifier [199]. In chapter 6 we give an example of NB as *indicator* nodes.

In the late 80's, with the addition of new concepts such as *d-separation* and new algorithms like message-passing [178] we are able to build more complex network structures. Now, graph theory is used to code variable's (in)dependence. The result is BNs models that can capture more complex interactions. As Pearl [177] comments

... the fundamental structure of human knowledge can be represented by dependency graphs and *that* mental tracing of links in these graphs are the basic steps in querying and updating that knowledge[3].

3.3 Graph Theory

The way variable's dependencies are determined depends on the type of graph: *direct* or *indirect*. We are going to focus on directed graphs given its relevancy from the point of view of expert elicitation. The use of undirected graphs are somehow constrained to the field of statistical modeling [170]. Note, however, that we can also derive a BN from an undirected graph; for more information on this later point see [62, 152].

To help the reader follow the discussion on the contribution of graph theory to BNs we are reproducing some definitions from the graph theory; these are taken from Castillo et al. [62].

Definition 1. ***Graph.*** *A graph* $\mathbb{G} = (X,L)$ *is defined by two sets* X *and* L*, where* X *is a finite set of nodes* $X{=}X_1,X_2,...,X_n$ *and* L *is a finite set of links, that is, a subset of all possible ordered pairs of distinct nodes.*

Definition 2. ***Directed Link.*** *Let* $\mathbb{G} = (X,L)$ *be a graph. When* $L_{i,j} \in L$ *and* $L_{j,i} \notin L$*, the link* $L_{i,j}$ *is called a directed link. A directed link between nodes* X_i *and* X_j *is denoted by* $X_i \rightarrow X_j$

Definition 3. ***Undirected Link.*** *Let* $\mathbb{G} = (X,L)$ *be a graph. When* $L_{i,j} \in L$ *and* $L_{j,i} \in L$*, the link between nodes* X_i *and* X_j *is called an undirected link. An undirected link between nodes* X_i *and* X_j *is denoted by* $X_i - X_j$ *or* $X_j - X_i$

Definition 4. ***Directed and undirected graphs.*** *A graph in which all the links are directed is called directed graph and a graph in which all the links are undirected is called undirected graph.* See Figure 3.2.

Definition 5. ***Path between two nodes.*** *A path from node* X_i *to node* X_j *is an ordered set of nodes* $(X_{i_1},...,X_{i_r})$*, starting in* $X_{i_1} = X_i$ *and ending in* $X_{i_r} = X_j$*, such that there is a link from* X_{i_k} *to* $X_{i_{k+1}}$*, k=1,...,r-1.*

The length of this path is (r-1), the number of links it contains.

[3]Italics added

Definition 6. ***Cycle.*** *A cycle is a closed directed path in a directed graph.*

Definition 7. ***Cyclic and acyclic graph.*** *A direct graph is said to be cyclic if it contains at least one cycle, otherwise it is acyclic (Cycle refers to a path where the initial and terminal node is the same and that does not use the same link more than once. A path is a sequence of consecutive links in a graph).*

Definition 8. ***Parents and children in directed graphs.*** *When there is a directed link $X_i \rightarrow X_j$ from X_i to X_j, then X_i is said to be a parent of X_j and X_j is said to be a child of X_i.* See Figure 3.2b.

(a) Undirected graph. Graph with undirected links. (b) Directed graph. Graph with directed links.

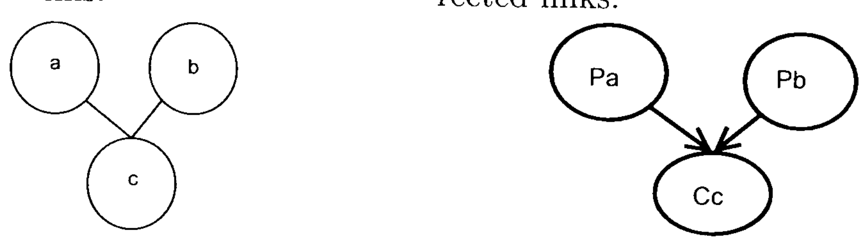


Figure 3.2: Nodes Pa and Pb are the *parents* of node Cc and node Cc is their *child.*

The nodes in a directed graph represent random variables, which can be discrete or continuous, and links denote the conditional dependency among the nodes. Thus the absence of links shows the absence of direct relationship between them. In the case of directed graphs, the node's dependency is given by the direction of the arrows, i.e. from parent to child[4].

Definition 9. ***Tree.*** *A connected undirected graph is said to be a tree if for every pair of nodes there exists a unique path.*

Definition 10. ***Simple tree and polytree.*** *A directed tree is called a simple tree if every node has at most one parent. Otherwise it is called a polytree.* See Figure 3.3a.

Definition 11. ***Trees and multiply connected graphs.*** *A connected directed graph is said to be a tree if the associated undirected graph is a tree; otherwise, it is said to be multiply connected.* See Figure 3.3b.

Naïve Bayes is an example of probabilistic model using tree graph. NB is constrained to one type of structure, see Figure 3.1. BN models, on the other hand, can be represented

[4]For undirected graphs, this dependency is present in the nodes probability tables.

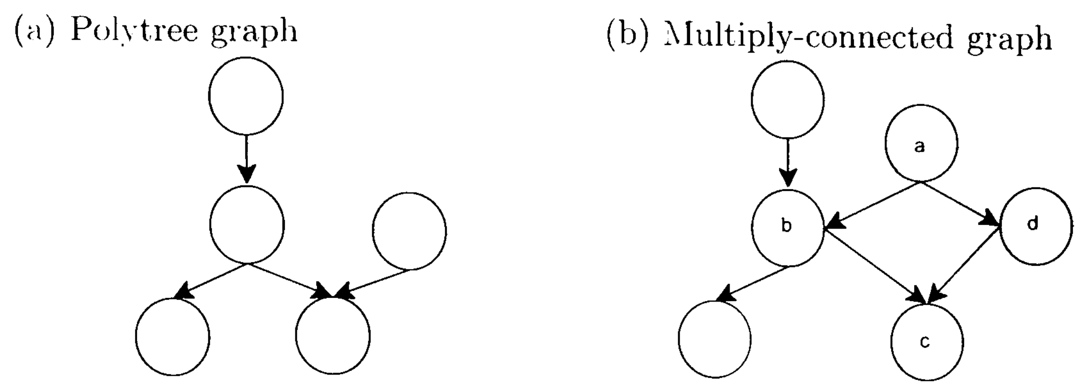


Figure 3.3: The difference between (a) and (b) is that the later contains a *bucle* formed by the nodes a, b, c, d.

using a variety of graphs (constrained only on the availability of algorithms to calculate them) see Figure 3.3.

The difference between NB and BN derives from the variable's conditional independence rules. On NB this independence is "axiomatic" while in BN it derives from graph theory and follows the concept of *d-separation*.

The next section reviews the concepts of independence, conditional dependence and d-separation which are central to the understanding of BNs.

3.4 Independence, conditional independence and directional-separation

Let us begin by providing some definitions, these are taken from Castillo et al. [62].

Definition 12. ***Independence.*** *Let X,Y be two disjoints sets of random variables, we say that they are independent, written as : $X \perp\!\!\!\perp Y$, if and only if:*

$$p(x|y) = p(x) \tag{3.6}$$

That is, the knowledge of Y does not influence, does not condition, our knowledge about X. In the case of disjoints sets of random variables $X_1,...X_n$ we can say they are independent if and only if:

$$p(x_1,...x_n) = \prod_i^n p(x_i) \tag{3.7}$$

that is, the combined probability is equal to the product of their marginals. The *marginal* *probability distribution* $p(x_i)$ is obtained by

$$p(x_i) = \sum_{x_1 \dots x_{i-1}, x_i \dots x_n} p(x_1, \dots. x_n)$$

Definition 13. *Conditional independence.* *Let X, Y, Z be three disjoint sets of random variables, we say X is conditionally independent of Y given Z, written as : $X \perp\!\!\!\perp Y|Z$, if and only if:*

$$p(x|z,y) = p(x|z) \quad or \quad p(x,y|z) = p(x|z)p(y|z) \tag{3.8}$$

Note also that if $X_1, ..X_n$ are conditionally independent given another set, $Y_1, ..Y_n$ then we have

$$p(x_1, ...x_n|y_1, ..y_n) = \prod_i^n p(x_i|y_1 ..y_n) \tag{3.9}$$

How can we represent these node's conditional (in)dependence in a graph? Let us study the concept of *d-separation* or (directional separation), first introduced by Pearl [178].

Definition 14. *d-separation.* *Let X, Y and Z be three disjoints subsets of nodes in a directed acyclic graph $\mathbb{G}$; then we say Z d-separate X and Y if and only if along the undirected path between any nodes in X or in Y there exist an intermediary node A such as*

- *A is a convergent node in the path and neither A nor its descendants are in Z, or*
- *A is not a convergent node in the path and A is in Z.*

Following this definition we can distinguish three types of connections in a direct graph:

Serial connection. The nodes $X \perp\!\!\!\perp Y|Z$, otherwise, they are dependent. That is, if we know about Z then, more information on X will not change our knowledge on Y; X becomes irrelevant to Y once we know about Z. See Figure 3.4. Note that the relation of independence does not change if both arrows change direction.

Divergent connection. The nodes $X \perp\!\!\!\perp Y|Z$, otherwise, they are dependent. In this case, X and Y share the same parent Z. If we know more about Y this will influence its

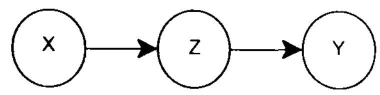


Figure 3.4: Serial connection

parent Z and in turn Z will influence its child X (and vice-versa from X to Y through Z). However, if we know Z then any additional information on either X or Y will not affect each other. See Figure 3.5.

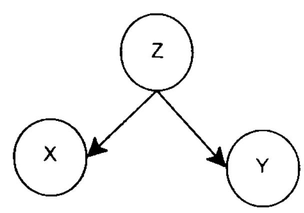


Figure 3.5: Divergent connection

Convergent connection. The nodes $X, Y|Z, Z_1$, otherwise, they are independent. In this case, X is independent a priory of Y and vice-versa $X, Y|\emptyset$. However, if we know Z or any of its descendants Z_1 then X and Y become conditionally dependent, so that, any information on X will affect Y and vice-versa. This is what is called *Inter-causal reasoning* or *explaining away*. That is, if Z is explained by X alone then the probability of Y will be reduced, given that both X, Y are the only two available explanations for Z. See Figure 3.6.

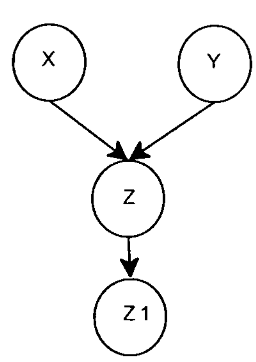


Figure 3.6: Convergent connection

P. Krause [170] points out

> ...no rule for detecting independence in a DAG can improve on d-separation in terms of completeness.

Figure 3.7 shows how two models based on two different assumptions of independence

(a): $P(A,B,C,D) = P(A)P(B\|A)P(C\|A)P(D\|B,C)$
(b): $P(A,B,C,D) = P(D)P(A\|D)P(B\|D)P(C\|D)$

Table 3.1: Joint probabilities.

can represent the same problem domain. Note the different joint probability distribution they produce as a consequence on Table 3.1.

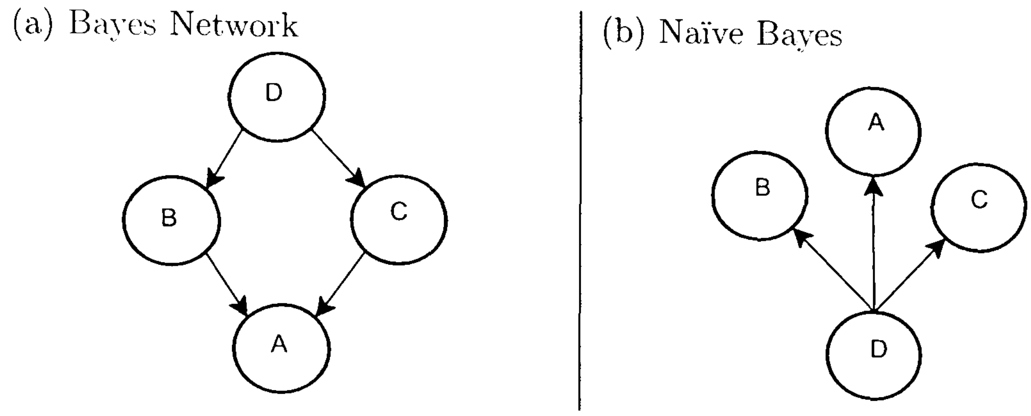


Figure 3.7: a) Multiply-connected graph and b) Constraint to be represented with only one type of structure where the node D is the target node.

We are now in a better position to understand how the AI application of Bayes theorem has evolved. On section 3.2 page 25, we commented how the joint distribution, see equation 3.4, conditioned all the variables on each other as opposed to the joint distributions on Table 3.1a) which uses the concept of d-separation. As Cooper [38] signals:

> A key advantage of a BN is that it represents probability relationships concisely. To build one, it is necessary to consider only the known dependencies among the variables in a domain, rather than to assume that all variables are dependent on all other variables. This provides an efficient and expressive language for acquiring and representing knowledge in many domains.

That is, the information relevant to a node is provided by the parent(s) which are conditioning its probability [80]. This dependency is used during elicitation, as Krause [170] comments

> The notion of relevance embodied in the use of conditional probabilities also influences the elicitation of the probability values.

It also helps building and understanding models as a concatenation of autonomous concepts, i.e modules. Laskey and Mahoney [198] signal that

> The ability to represent conceptually meaningful groupings of variables and their interrelationships facilitates both knowledge elicitation and knowledge base maintenance

In the context of building BN models, the concept of (in)dependency, explicitly presented in the graph, is of particular interest if one thinks that assuming incorrect dependencies, on its own, can affect the overall model's output even more than introducing wrong probability estimates [74].

3.5 Bayes Network

A BN is an acyclic directed graph[5] with its associated joint probability distribution that follows the rules of *d-separation*. The nodes are quantified using data or expert's judgments or both and represent the strength of the relationship [62,111,152,178]. Note also that this asymmetric relationship between parent and child nodes can be read as a cause $\rightarrow$ effect relationship; which from the point of view of elicitation has a number of advantages. Section 3.6 expands on this latter point.

The notion of a node's dependency has several consequences:

- As a directed graph the arrows show the direction of the dependency or in their absence conditional independence.
- Computational tractability. It makes possible the inference of reasonable sized networks. When variables are independent their joint distribution is just the product of their marginals. There are known exact algorithms, e.g. message-passing [176] or polytree algorithm [178] that can perform probabilistic inference using singly connected networks in time that is linear as a function of the size of the belief network or in case of multiple-connected graphs by transforming the network into a singly connected one using a clustering algorithm[6] [111,129] or using approximation

[5]If the graph was cyclic, e.g. $A \leftrightarrow B$, the evidence on A would affect B and this new evidence on B would have an impact on A,thus, following an endless cycle.

[6]This is because multiply connected graphs contain at least one *bucle*. Hence, the computation becomes NP-Hard as proved by Cooper [38]

methods [62]. The clustering or joint tree algorithm is the one most often used to calculate BNs.

- Locality. We assume that a node is independent of the rest of the nodes in the network given its parents following Markov conditioning, thus equation 3.4 on page 25 becomes equation 3.10.

$$p(x_1,..x_i,..x_n) = \prod_{i}^{n} p(x_i|pa(x_i)) \tag{3.10}$$

where n stands for the number of parent nodes pa of x.

- Modularity. A network can be explained as a joint set of several modules, each one explaining part of the model.

From these properties we can derive a number of conclusions:

- Computations can be done locally. This locality can be seen as groupings of inter-related information captured in a network module.

- Modularity helps us build the model following a *divide and conquer* approach. We can reason about parts of the model independently of the rest of the model. In this fashion, BN models can state very complicated hypotheses composed from simple elements.

  - Each part of the network is self-explained by the task they represent and by the causal relations that join them as Mahoney and Laskey [198] comment

    > ... decomposition must be both semantically separable and formally separable. Semantic separability means that the sub-problems into which the problem is decomposed are meaningful to the expert and posed at a natural level of detail. Formal separability means that the subproblems are capable of being re-aggregated into a complete and consistent probability model.

  - BN's modularity could be used to break the problem down into smaller sub-processes, each sufficiently distinct to help the expert provide information [198]. As Pearl [177] comments

> Human performance ... exhibits a different complexity ordering: probabilistic judgments on a small number of propositions ... are issued swiftly and reliably, while judging the likelihood of a conjunction of many propositions entails a great degree of difficulty and hesitancy. This suggests that the elementary building blocks which make up human knowledge are not the entries of a joint-distribution table but, rather, the low-order marginal and conditional probabilities defined over small clusters of propositions.

- Modules can be interpreted as classes following a Object Oriented Bayesian Network approach (OOBN) [19, 122, 158];
- Following an OOBN Fenton and Neil [137] developed a set of heuristics they called *idioms*. They are a small number of natural and reusable reasoning patterns. The knowledge engineer simply compares their current problem, as described by the expert, with the idioms and reuse the appropriate one for the job. These idioms correspond to a concrete problem that can be captured in a small sub-net. These are then joined to create a large BN. This technique has been validated in numerous projects and has been formally incorporated into BN tools such as Hugin [13] and AgenaRisk [134].
- Sensitivity analysis can be conducted exploiting the dependency among the nodes in the path [40, 41].
- It can also help develop BNs Ontologies. In chapter 4 section 4.3, we explain how the development of new ontologies can offer an unifying methodology to building BN models.

3.6 Causality and BNs

Relationships among nodes can be interpreted from a probabilistic point of view, i.e. as conditional dependencies or from a subjective point of view as causally related [212].

The difference between probabilistic and causal interpretation is the same as between *shallow* knowledge based on correlation and *deep* knowledge based on causality. This is because using causality to build a model entails the knowledge (the structure of the variables' relationship) of the problem domain [212] and as Pearl [179–181] argues this deep knowledge regarding causation:

- leaves little ambiguity between the knowledge engineer and the expert about what is being said;
- guides the knowledge engineer determining which analysis to conduct;
- experts just have to follow the model's causal flow to explain it and
- consolidates the validity of the model.

We agree with Druzdzel and Simon [59] when they comment that although causality cannot always be asserted it is always useful as a tool to elicit and understand BN models

> While it is certainly not the case that every directed arc in a BBN denotes causality, the formalism is capable of representing asymmetry among variables, and thereby, causality.

Argument that it is also shared by Lauritzen and Spiegelhalter [129]

> "causality" has a broad interpretation as any natural ordering in which knowledge of a parent influences opinion concerning a child. This influence could be logical, physical, temporal or simply conceptual in that it may be most appropriate to think of the probability of children given parents.

According to Pearl [181] to find causal relations we must

> ...identify the clues that prompt people to perceive causal relationships, given that statistical analysis is driven by covariation, not causation, and assuming that most human knowledge derives from statistical observations.

These clues can help gaining knowledge to elicit the model:

- Temporal. The probabilities referring to the present time are *conditioned* only on what happened earlier, not on what happens later. Pearl's [178] sees a temporal feature as a more intuitive way to explain the model rather than a property of the model itself.

- Contextual. Causal networks have a context where the relations hold: outside this context the model fails. As a result of this, causal models cannot be applied to other problems by "analogy" unless they share the same context.

  Note that in Bayesian probability all probabilities are conditional, i.e. they are contextual. The notation $P(A)$ for the probability of A is based on the evidence $P(A|\xi)$, where ξ is the context where the estimation is taking place. In the case of eliciting a BN model it represents the expert knowledge about the domain. This contextual measurement is always taken into account, e.g. $P(A|B,\xi)$ and thus not always written, $P(A|B)$ [36].

- Spurious correlations can disappear given that causality assumes a deeper domain knowledge that is able to explain this causal relation.

- Asymmetric. The causal link goes only from parent to child. This assertion is consistent with causal reasoning. For instance, we can assume that the "financial auditing" is *caused* by "internal fraud". An increase in frauds is followed by an increase in auditing. However, this is a wrong dependency assumption, i.e. "internal fraud" is *caused* by poor auditing. The lack of auditing increases the likelihood of frauds and not the other way around.

- Manipulation. Two events are causally related when we manipulate one and see a change in the other. The relevance of this change is what makes two events to be causally correlated.

- Weak transitivity. For instance, in the case of a serial connection if an event A has an impact on event B and event B has an impact on event C we conclude that A has an impact on event C. However, this transitivity is conditioned upon B and as we discuss in section 3.4 this dependency can be broken if B is known[7]. Thus the "weakness" of this transitive relation[8].

[7]Note that this also applies to other types of connections where the path between two variables is d-separated.

[8]There is a caveats on this point: It is also true that this weak transitivity may not hold when two events A and C are simply correlated. In this case, these two events may have a hidden parent node that makes them dependent in spite of knowing B. However, as Giarratano observed [81]

> .. the fact that this assumption is not followed does not affect the overall result when we are interested more on the general behaviour of the system rather than individual numerical results.

On building large-scale BN models Druzdzel et al [57,59] found that in practice it is very useful to interpret BN models as causal models for the following reasons:

- easier for the expert to understand and conceptualise;
- facilitate the interactions among multiple experts inasmuch as causality provides a common language among the experts;
- usually it ensures satisfaction of the Markov condition (variable independence is explicitly shown in the graph [152]);
- usually it is easier to obtain probability judgments following the causal direction.

From the point of view of explaining the models, Lacave and Díez [128] add other reasons:

- the identification of *invariant* causal relationships in a domain allows the prediction of effects of both spontaneous causes and actions;
- the concept of explanation is very closely tied to the notion of causation.

3.7 Summary

We have explained how Bayes Theorem can condition domain uncertainties on observable events, thus allowing us to express in probabilistic terms rare events such as mid-air aircraft collision or financial breakdowns.

The use of BN enables us to capture complex networks whose joint probabilities, explicitly represented in a graph, are the result of applying the concepts of (in)dependency and directional separation. Applying these concepts, we can think of a BN model as the addition of different modules and this makes possible the representation of real-world problems.

A BN model can be based on quantitative or qualitative knowledge, or both. In the case of qualitative knowledge, probability theory, as we discuss in the following chapter, grants consistency to the elicited knowledge base. Note that the elicitation of a model's relevant variables are given by the variables dependencies. This dependency can also be interpreted as cause and effect, thus providing a rich language to elicit the model's relations.

The challenge is to elicit the knowledge base that informs the BN model, its topology and its NPTs. The following chapter focuses on eliciting NPTs, it documents the elicitation problems and reviews current methods.

Part II

Modelling and Eliciting Node Probability Tables

Chapter 4

Modelling and Eliciting NPTs

4.1 Introduction

There are two approaches for modelling BNs, namely Data Based and Knowledge Based:

- Data based approaches: This methodology makes use of algorithms to learn BNs from known (training) data [62, 146, 152]. The assumption is that there is a hidden pattern in the data that these algorithms can discover. There are different algorithms available that search through the database for possible networks. They can be grouped into two main categories [88]:
  - Constraint-based search: These algorithms determine a network describing the data that satisfies a given constraint. For instance, in Pearl's Inductive Causation (IC) algorithm the constraint is that the model structure must follow a *causal chain* [178];
  - Bayesian search: Constructs the graphs with the highest posterior probability given the data. This approach is guided by the actual *correlation* of the given data [38].

In any case, more often than not, models coming out of learning algorithms need to be fine-tuned by an expert. The addition of expert knowledge can result in a model reflects the nature of the data better than the outcome produced by searching algorithms alone [149].

- Knowledge based approach (KBA): A model is seen as a cognitive map of an expert's knowledge and experience. In the context of BNs, we interpret these cognitive maps as causal maps that express the judgment that certain causes will lead to particular effects [150]. As Russel and Norving [197] comment

  > If the knowledge-base is true in the real world then any sentence derived from it by sound inference must be also true in the real world.

The domains of interest for this thesis, i.e. high safety risk and operational risk in financial institutions, are characterised by the absence of data and the availability of expert knowledge. Hence, this thesis focuses on the KBA.

Expert knowledge together with the semantic characteristics of BNs will enable us to model these domains accurately and effectively[1].

The need to elicit expert knowledge has hindered the development of BN models and it arguably remains a significant bottle-neck in the application of BNs.

Indeed, the literature reports general guidelines on how to bring about the elicitation process [17, 34, 144]. All of these agree, in general, that the elicitation process should be carried out in a number of broadly defined steps [76]: expert selection, definition of the scope of the problem, selection of the granularity of the description, identification of the relevant factors, structuring, iteration, encoding, verification, documentation. The differences between the various methods lie in the greater or lesser emphasis placed on some of the steps.

Regarding BNs, however, we have not found a unifying elicitation methodology as such. In most cases, elicitation methods are tied up to the type of problem domain (e.g. medicine, financial risk), to the availability of experts and their expertise, the domain novelty and the type of information the experts must provide (network topology or estimates of probabilities).

[1] We are referring to BNs as a *semantic* approach as opposed to *syntactic* approaches like Rule base. The latter follows classical logic to define rules, i.e. *if ... then*, and the variables' scope, e.g. $\forall x$, $\exists x$. BN-style semantic approaches, on the contrary, employ conditional probability to define rules, i.e. *B given A*, and use semantics to define the variables' scope, e.g. by representing a BN as a causal model.

See Krause and Clark [124] for a clear, albeit brief, differentiation between the two approaches, and Pearl [178] for details on concepts such as intension and extension.

Other sources of variability involve the form in which experts' answers are given, e.g. using causality to explain model's behaviour, and of course budget/time constraints[2] [22, 26, 225].

This lack of an unifying methodology explains the variety of techniques available. In section 4.4 we review these techniques in the context of eliciting qualitative expert knowledge after having review in the next section the context in which these techniques are thought out.

4.2 Subjective elicitation - Heuristics and Biases

There is a long-standing discussion on the legitimacy of the subjective approach to building models based on uncertainty [27, 109, 189, 200]. The main criticism [110] on the use of subjective probability is that, according to De Finetti [52] and Savage [200], for an estimation to be valid it only needs to be consistent with the expert's judgment. However, as the work of Kahneman et al.'s [5] showed, in the case of a systematic bias,

> ... the judgment can be consistent with the expert's rational but not according to the probability of the event. Thus, for subjective probabilities to be considered adequate, or rational, internal consistency is not enough.

The aim of Kahneman et al.'s [5] work was to produce a "normative" model of how humans act when facing uncertainty. Their research showed that conditioned reasoning based on heuristic rules sometimes produces unexpected outcomes. As Kahneman et al. [5] comment:

> People rely on a limited number of heuristic principles which reduce the complex task of assessing probabilities and predicting values to simpler judgmental operations. In general they are quite useful but sometimes they lead to systematic errors.

Such heuristic-induced errors are known as *heuristic bias*. In the following we briefly review some of the forms in which this may occur [5], as illustrated by the diagram in Figure 4.1.

[2] For further reading on expert elicitation: Ayyub [17] provides a introduction to the subject. Meyer's book [144] is mainly based on the findings of the elicitations carried out by the USA program for Nuclear Waste [168]; it provides a good example of what is expected of knowledge elicitation and what its limitations are.

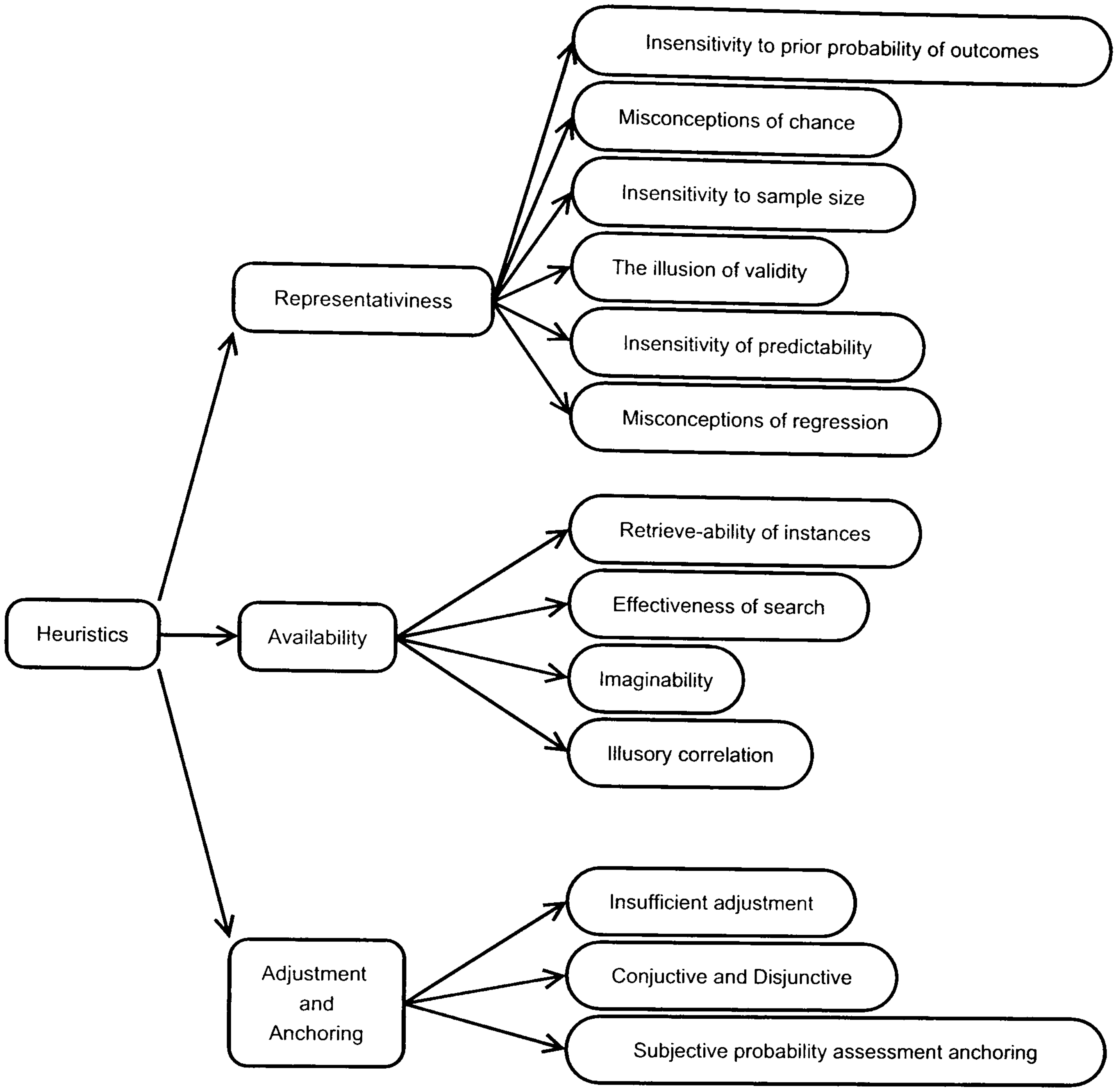


Figure 4.1: Sources of heuristic bias according to [5].

Representativeness bias refers to people evaluating the probability of an event based on how similar it is to a general class. For example, how likely I am consider a person to be a lecturer depends on my pre-conceptions of how lecturers should be. *Availability bias* derives from the tendency to assess the probability of an event based on one's recollection of a similar event. For example, I assess the risk of heart attack based on how many people I can recall have had a heart attack. *Adjustment and Anchoring* occurs when people's initial estimates are not adjusted sufficiently to reflect the actual event, resulting in an overconfidence in their estimation which is not always backed by their knowledge.

In summary, Kahneman et al.'s [5] research concludes that biases are an inherent part of human cognition. The fact that sometimes heuristic bias may lead to errors has put into question the coherence and calibration[3].

As Ayton and Pascoe [16] argue:

> ... if human decisions are plagued with unconscious biases, how might biases in the knowledge base be avoided?

For this reason A. O'Hagan [155] points out most research goes toward the use of

> ... improper, so-called non-informative priors ... *given* ... the illusory objectivity of these priors ... *however* ... they must **never** be used without first thinking whether genuine and informative prior knowledge exists[4].

Let us review the grounds for O'Hagan's assertion. From the point of view of a BN we can say that, inasmuch as it follows the laws of probability, its coherence is granted as long as Cox's rules [140] are held.

P. Krause and D. Clark [124] comment that one of the reasons for using probability theory, and in particular Bayes reasoning, to develop expert systems, is to avoid the biases of other less *rigorous* approaches.

We have two examples of this (note that these two approaches are not longer in use):

- The Certainty Factors approach could not grant the coherence of the elicited values given its *ad hoc* approach to assigning numerical values to events. It must be said that this approach derived from the type of domain it modeled: medicine. Shortliffe [206] explained that medical experts agreed to provide their belief in a given conclusion, but that this belief *in favour* of the conclusion should not be constructed as evidence *against* it. In words of Harré[5]

  > ... to confirm something to ever so slight a degree is not to disconfirm it at all, since the favourable evidence for some hypothesis gives not support whatever to the contrary supposition in many cases.

[3]Note that calibration refers to the expert's knowledge about the "true" nature of the event and/or to the accuracy of the elicitation method used to extract this knowledge.

Elicitation is only concerned with the latter point. Indeed, it is against experts' opinion that the model is going to be tested and used.

Nevertheless, there is an interesting discussion on this detachment between expert's knowledge and the actual nature of the event. For D.R.Cox [116] this detachment from the quality of expertise ... *seems a bad idea likely to perpetuate the errors of the past* while for Kadane et al. [116] thei detachment is what make possible models based on expert's opinions ... *there is not future errors but actual knowledge.*

[4]Italics added

[5]Quoted in [206]

Shortliffe [206] uses the example of diagnosing an infectious disease due to streptococci from its associated chain of symptoms. He explains that while experts provided a 70% belief on the conclusion given the symptoms they felt uneasy to state the correspondent 30% against it as we would do if we were to follow probability theory: $P(A) + P(\neg A) = 1$.

Note the expert is not irrational, as we would conclude if we were to follow probabilistic reasoning; rather, as Buchanan and Shortliffe [24] would say, he is only *reflecting a level of belief*.

To support this argument Shortliffe followed the work of R. Carnap [27] who interprets *degrees of confidence* as *degrees of confirmation*, inasmuch as the probability of a statement is the degree of confirmation that the empirical evidence gives to the statement [91]; that is, for Shortliffe [206] the

> ... term confirmation does not indicate that an hypothesis is proven but rather that an observations lends credence to it.

That is why Shortliffe did not use the formalism of Probability Theory and based his ES MYCIN on the Confirmation Theory. For him the domain of Medicine did not have the level of knowledge assumed in probability theory; thus the creation of the Certainty Factor approach based on ad hoc rules made for the specific domain of Medicine[6]. Note that this lack of a formal theory behind this approach is reflected in the inconsistency of some of its conclusions.

- The project PROSPECTOR, used to find mineral deposits, is another example. Although this project is sometimes quoted as an example of a Bayesian model, this is not truly the case in reality. During the development of this project experts showed inconsistencies in the elicited value. For that reason the results of linear interpolation were used to assure the consistency of the elicited values [81].

In conclusion, regarding coherence, we can interpret probability theory as the normative standard of how people should make judgments facing uncertainty [16, 60, 170].

[6]See Giarratano [81] for an introduction to Certainty Factors and E. H. Shortliffe [206] for a more detailed explanation.

This leaves us with the problem of how well expert's elicited assessments adjust to the "true" nature of the event, i.e. how well calibrated they are. On this area, recent research has questioned the grounds for Kahneman et al.'s [5] assertions.

The current debate is more about research on improving elicitation methods than about questioning the value of expert's opinion. As G. Wright and P. Ayton [76] discuss

> We should remember though that an expert excels in his ability by knowing "what to look for" and not in the ability to integrate information. It is in this aspect where an expert judgement can be improved on.

For Hoffrage [75, 101] heuristic bias has more to do with a lack of methods to elicit the experts' opinion than a problem with the nature of human cognition. As Ayton [16] points out

> ... the experimenter's assumption he is able to calculate what the subjective probabilities *should* be for all of the subjects was absolutely necessary if one was to judge judgment. However, it is also an indication of the artificiality of the task.

The experimenter is treating subjective probabilities as objective ones. Gigerenzer [82] argues that if people provide similar answers it is because they are able to calculate them. However, in doing so the elicitation exercise is no longer subjective [125].

For Chapman [133] the current debate is more about

> ... the standards used to judge judgment rather than the judgments themselves.

Under this respect, human bias is seen as the result of ill conceived tasks or of the use of a wrong normative model. Ayton and Fischer [15] raise an interesting point by questioning the heuristics of "representativeness". They compare the "gambler's fallacy" with the "hot-hand fallacy". Both are different interpretation of the same event: a gambler betting following short or long term frequencies of the draws, respectively. Both interpretations are wrong if we think that draws are random events. Hence, they raise the question of how humans must interpret the result: do humans misunderstand the concept of randomness? If we do, as these two fallacies seem to show, how can we measure it against a statistical concept? Which in turn begs the question: do we know the human reasoning process well enough to base conclusions on experiments that may fail to capture that process? Are we applying the right "normative" model?

Gigerenzer's research [82] has shown that biases such as "Availability" disappear when subjects are confronted with frequencies, hence concluding that some biases are the result of how information is presented. Base-neglect [16, 229] offers an example in point. Base-neglect happens when people ignore prior probabilities and base their judgments entirely on facts that may not be relevant to the problem at hand. Take for instance the fear of dying in a plane crash. There is indeed a great probability of not surviving a crash: however, it is very unlikely to have one in the first place. The fact that the experimenter expects the subject will think following Bayesian reasoning does not imply that they will; for instance, by changing the way the information is presented the subject may not need prior probability [121]. Garthwaite et al. [79] give another example regarding "Overconfidence bias" when eliciting probability estimates for compound events, this bias is reduced when events are elicited separately rather than combined.

There is also a discussion on whether single events can be assessed [16]:

> ...no statement about confidence in a single event can violate any laws of probability ... *hence if we cannot compare two judgments then* ... who can say when an error has occurred ...[7]

This questions whether single probability assessments may be the product of lab experiments that cannot be reproduced in a real situation. G. Wright and P. Ayton [229] comment on the example of business entrepreneurs who are asked about the probability of success of their businesses. The experiment concludes that participants showed a great deal of "overconfidence". However this can be also explained as the difficulty for the participant to forecast without taking into account a number of variables:

> A business does not have a probability of going well or bankruptcy but is the parts that is compose with that should be assessed, this is what validates scenario planning.

Along the same line, Krueger's [125] research concludes human biases derive from unrealistic expectations tested by methods that ignore fundamental notions of human

[7]Italics added

cognition. On this line, Fischoff [72] provides a number of reasons for human biases, e.g. "subjects being unable to express what they know", "unfair tasks, misunderstanding the task they must perform" or "maybe expecting too much of the expert's knowledge", "deficiency from the experimenter to understand the respondent's conceptual universe", among others.

Bolger [22] points out that human bias, in many instances, lacks *ecological validity*[8] and when found does not necessarily mean a bad performance. For Hamm [93], research on human biases and heuristics should be used to highlight the limitations of the subjects' knowledge and to develop methods to alleviate them [60].

Budescu and Karelitz [216] explain bias as a communication problem between expert and knowledge engineer. They distinguish three modes of communication: *Numerical, Range and Verbal*, ranging from the precise to the ambiguous. For them, problems during elicitation derive from miscommunication among modes, e.g. the knowledge engineer prefers the Numerical mode while the expert prefers the Verbal mode. This asymmetry between communication modes is the origin of the misunderstanding.

Budescu and Karelitz accordingly devise a translation method that associates a function to each mode. Their aim is to reduce the effects of variability among the participants and to provide the most accurate representation of the experts' opinion.

For Ayton and Pascoe [16] these findings, in general,

> ... justified exploring and developing methods for representing uncertainty in expert knowledge. If human judgement and knowledge about uncertainty are more meaningful and more exploitable when elicited and modelled in the appropriate fashion ...

In the case of BNs, we are able to test the experts' calibration; Lauritzen and Spiegelhalter [129] comment that the experts' calibration is confirmed as more feedback comes along:

> ... consistent "surprise" for a particular case would indicate an "outlier", while over a number of cases, build-up of 'surprise' in a particular part of the system may indicate either faults in the numerical assessments or the

[8] The degree to which the laboratory tasks correspond to what the experts really do, and therefore experimental findings can be generalized to the real world.

structuring.

An example in point is the weather forecast ES "Hailfinder". A. Murphy and R. Winkler [18, 87] show how the experts modelling weather forecast benefit from frequent feedback. The MUNIN project provides another example. This project was developed to assist in the diagnosis of neuro-muscular disorders [124]. P Krause and D Clark [124] explain how initial probability estimations are strengthened when the system accumulates data. This is achieved by means of the method used to build the probability tables, explained in section 4.4.

When data and/or feedback are not available the use of sensitivity analysis can help refine the model's output [103]. As van der Gaag and V. Coupé [218] comment

> The basic idea of performing a sensitivity analysis is to systematically vary the assessments for the network's conditional probabilities ... *some* ... will show a considerable impact, while others will hardly reveal any influence[9].

This goes to show that the development of a model is an iterative process where model knowledge is fine-tuned with the help of sensitivity analysis and/or as new data comes along.

Also BNs interpreted as causal models entail a deeper knowledge about the domain, with the ability to explain its internal relationships and to assist the experts' judgment by providing the relevant variables to consider.

Developing a unifying methodology would reduce the bottleneck of eliciting BN models by creating and maintaining real problem domains using a standard approach that can be readily validated. Ontologies may be the answer to this question.

4.3 Ontologies

The aim of using ontologies to develop BN models is in part to answer the need of a unifying approach to building models and validating their outcome. The benefit for BNs is that information can be maintained, discussed, augmented, re-used and documented.

For V. Alexiev [9]

[9] Italics added

> Ontologies typically define concepts and their relationship, together with constraints on those objects and relations.

An Ontology allows the experts to translate their "mental models" into explicit and re-usable models that would be general and at the same time independent of the problem domain [203].

The aim is to provide a shared and documented understanding of a domain that can be communicated between people [68]. To achieve this generality we need a level of abstraction, i.e. meta-information, that describes the domain away from any specific detail [48].

To validate these abstract descriptions we need to develop an *objective* standard. As Gruber [86] comments

> Formal Ontologies are viewed as designed artifacts, formulated for specific purposes and evaluated against objective design criteria.

The design criteria proposed by Gruber [86] include[10]:

- Clarity: definitions should be objective, in the sense that they should be independent of the social or computational context. Whenever possible they should be captured using logical axioms.
- Coherence: any rational user should be able to infer the same conclusions according to the definitions stated.
- Extendability: an ontology should be general enough (by limiting the number of assertions) so that further specialised ontologies can be derived from it *monotonically*. In this sense building a model could be seen as putting together different ontologies. We can identify two approaches [9]:
  - merging data models to create a single ontology or
  - mapping existing models. In this case, the ontologies remain unaltered but for the addition of a number of links between them.

[10]These criteria act as guidelines. We can find other approaches that, based on the same criteria, expand or reduce the definition of ontology depending on the domain. See [4] for an example of building an ontology on the domain of Biology

- Minimal encoding bias: the conceptualisation of knowledge should be independent on the tool (e.g. elicitation method) used to capture it. That is, we should not be partial to the most practical or familiar method, but rather choose the one that should provide the same results when encoded in different tools (e.g. using different elicitation methods).

For Gruber [86] the extent to which these criteria can be applied depends on the knowledge available and on the domain in which the ontology is going to be used. For instance, in the engineering domain [230], characterised by an axiomatic knowledge, the resulting ontologies will better fit these criteria.

However, in domains characterized by a greater uncertainty we cannot always follow this criteria. In a BN coherence is granted by the use of probability theory, but we cannot say the same thing about the nodes' definition or topology. For instance, we may need to add some constraints in order to define concepts, e.g. in BNs the assumption of variable (in)dependence may not be shared by other experts. Thus we can no longer claim ontology extendability since other people may not share the same constraints. This falls short of the level of generality required to extend the ontology as we add more assertions about the model.

Regarding encoding bias, Helsper and van der Gaag [98] discuss the use of probability theory to encode a domain's knowledge:

> ...since a probabilistic network in essence is a model of a joint probability distribution, multi-valued domain concepts must be modelled as statistical variables, which are single-valued by definition.

Costa and Laskey [39] propose a probabilistic ontology to represent uncertain domains. The idea behind it is that although experts may have imperfect knowledge that falls short of the definition of a concept, they may still have some knowledge about it; the extent of that knowledge can be represented in probabilistic terms. For Costa and Laskey the knowledge about a domain application must include: types of entities in the domain, their properties and relationships, processes and events, statistical regularities and uncertainty regarding the domain among others.

They discuss the use of a new probabilistic ontological language, namely PR-OWL[11],

[11] The OWL language is the *de facto* standard used to develop web-ontologies. OWL is based

based on Multi-Entity BN (MEBN) first order Bayesian logic. MEBN [48] represents the world as made of entities that have properties and relationships. These entities' attributes and relationships are represented as a collection of fragments (MFrags) organized into Theories (MTheory). Each fragment specifies a conditional probability distribution given its parent fragment. An MTheory will represent the joint probability distribution that results from the combination of all the fragments.

For an example of this approach see the project "Detection Threat Behavior" [171]. In it, an MEBN structure is used to control the access of people to sensitive information. Information is available to all the users, but only the ones with privileges would be able to access it. Access is controlled using additional MFrags to refine the node's functionality.

See also Helsper and van der Gaag [97] for a didactic example on building a BN ontology. Their ontology focuses on the topology of the model.

These ontologies are developed using the tool created by the Standford University: "Protege"[12]. The University of Manchester and Amsterdam also developed the Ontology Inference Layer: "OIL"[13].

4.4 Probability Elicitation methods

Garthwaite et al [172] define the elicitation of NPTs as

> ... the process of formulating a person's knowledge and beliefs about one or more uncertain quantities into a (joint) probability distribution for those quantities.

Those quantities are elicited within the context of the node's conditional dependency. As we explained in Chapter 3, a node is independent of the rest of the network given its parents. Therefore, the number of parameters of a child node depends on the number of its parents and on the number of states they both have. For instance, for a node with five parents, assuming parent and child nodes are binary variables, we need to elicit $2^{5+1} = 64$ probability values, i.e. the number of values is an exponential function on the number of parents. For this reason we need techniques that make handling such a large amount of

on the XML language; without going into details it will suffice to say that this type of language provides meta-information about a concept.

[12] http://protege.stanford.edu/

[13] http://www.ontoknowledge.org/oil/

information possible.

This section reviews known methods to elicit the experts' opinion. The level of complexity varies; some methods are better suited to certain domains than others. e.g. Noisy-OR in Medicine; in general they address the problem of bias from different perspectives and their accuracy and the time taken to elicit may vary. These methods are:

- Interviews and Questionnaires.
- Verbal/Numerical scales.
- Probability wheels.
- Betting.
- Graphical tools: Bar graph, Probability wheel.
- Statistical distributions.

4.4.1 Interviews and Questionnaires

This method is of particular relevance for the present thesis. An example of its use can be found in Chapter 6, where we model the contribution of socio-technical factors to mid-air aircraft collision.

Researchers at the Safety Research Unit (SRU), in the School of Psychology of the University of Liverpool, lead by Prof. Ian Donald produced a questionnaire to find out employees' perception of management involvement in the risk management process [138]. This questionnaire is divided into 12 factors. For each of these factors a number of questions are used to defined a firm's risk perception. In Table 4.1, we observe how questions are grouped based on the factor they are more correlated with. In this case, factor one, "Personal evaluation of safety system". The last column of Table 4.1 shows the correlation between factor one and the answer to each of these questions.

As the NATS project's results have shown, questionnaires are a valuable tool to reach a wider audience, in this case over 7211 respondants, using the same set of questions to act as a normalising factor.

Questionnaires highlight the problems with direct elicitation methods: experts are more exposed to cognitive bias such as over-confidence, anchoring and adjustment. Gigerenzer [82] argues that direct assessments will not produce calibrated assessments "... *not*

Q. Num.	Question	Correlation
Q12	I feel satisfied with the safety information I get	.644
Q13	I am happy with the existing safety precautions for particularly hazardous work	.567
Q14	I feel satisfied with the attention given to safety in any training I have had	.556
Q10	I am happy with the safety equipment specified for my job	.549
Q9	Generally I am happy with the safety in my asset area	.483
Q22	I know the results of safety inspections to do with my job	.422
Q15	The people I work with are satisfied with the attention given to safety in any training	.420
Q8	The people I work with are satisfied with the information they get about safe working	.388

Table 4.1: Questions Correlated to Factor One: Personal evaluation of safety system. This is one out of 12 factors that have shown to define employees' perception of management involvement in the risk management process.

even offering French champagne" [82]. It is, thus, not surprising to see why the use of elicitation methods such as Delphi are highly criticised [17]. Questionnaires are also time consuming from the point of view of persons who administer the questions to the subjects who fill them [219]. In the case of Prof. Ian Donald's team it took several years to collect and to process the information. These amounts of time and the corresponding budget are not always available to develop a model.

4.4.2 Verbal - Numerical

Techniques like verbal scales try to address the problem of direct elicitation [92, 107, 193, 202]. The use of linguistic probabilities is based on the idea that humans feel more at ease providing probability estimations using verbal expressions, e.g. "Probable", than using numbers, e.g. "0.85" [217, 219, 225]. Renooij [193] recommends to couple verbal expressions with numerical values, in this case, verbal expressions are later evaluated to a numerical value.

However, the interpretation of verbal uncertainty may be influenced by cognitive biases [1]; there is also the difficulty of mapping numerical values to verbal expressions open to the different expert's interpretation [66], without mentioning the time it may take to elicit a complete NPT through this method.

However, it is also advisable to use techniques familiar to the expert (e.g. verbal scales) when the values to be elicited or the experts do not satisfy certain conditions.

Bolger [22] mentions the following:

- availability of repeated events of the same type;
- unambiguous occurrence of events;
- usable feedback and
- experience using probabilities.

As an example we can refer to our NATS project explained in Chapter 5. In it, we used this technique to obtain the expert's confidence about his assessment. The main reasons for using this approach was the scarcity of aircraft "near-misses", i.e. data, and the lack of statistical knowledge of the experts. We also share the view of Ngwenyama and Bryson [115] when they comment

> ... the use of conditional probability can help reduce the inherent problems of verbal-numerical probability elicitation by providing the expert with more information when assessing the probability ...

We provided a look up table where the experts could map their confidence using a verbal/numerical scale. The variance elicited was calculated and the results discussed with the experts, who could see the impact of their estimation on the model. Providing this quick "feedback" enabled us to discuss any possible misunderstanding or ambiguity in the values elicited; any disagreement between the elicited estimations and future data on near-misses will end up confirming or modifying these values.

4.4.3 Probability Wheel

A probability wheel, see Figure 4.2, has two sections of different colours. The size of the shadowed area is adjusted by the expert until the probability of the pointer stopping on the shadowed area agrees with the expert's estimation.

The idea is that when the pointer is spun the probability that it will land within the shadow area is equal to the probability assigned by the expert to some specified event.

4.4.4 Betting

The basic idea is that the expert is presented with the choice of two lotteries, schematized in Figure 4.3. He/she can either enter a lottery, where the pay-off is 10 if the expert

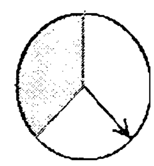


Figure 4.2: Probability wheel.

guesses the "correct" probability of a particular event, or accept a fixed amount of money that is offered instead. The idea is that the amount of money varies with "x" (see Figure 4.3) until the expert is indifferent about either bet.

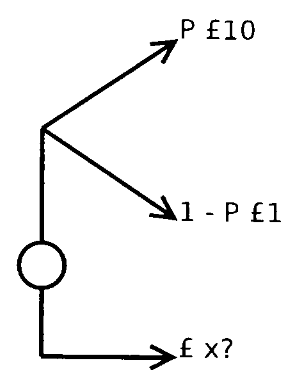


Figure 4.3: Betting method.

A similar approach is used in the field of finance where experts are paid with bonuses in relation to their forecasting. This practice has been criticised as experts may feel tempted to be *over-confident* in their estimations given the prospect of personal gains, as recent financial scandals have shown [67, 194].

Regarding graphical tools, Druzdzel et al.'s [88] research showed they are useful in elicitating experts' beliefs. They devised a game of cat and two mouse to evaluate how people learn probability values. The goal of this game is for the cat to catch the mouse. There are certain "conditions" on this game that the player must learn by trial and error. There are two main variables, one is the player's move (which is known to the player) and the other is the state of the game which is uncertain to him. If the player wants to catch the mouse he will have to learn, condition, his movement to that of the mouse. Participants were asked to produce probability estimations following three different methods: directly asking for numerical values, a probability wheel and a scaled probability bar.

Out of these methods, the scale bar was the most efficient in terms of accuracy and time spent recording the probability values. Whitcomb et al. [226] evaluated similar elicitation methods: numerical probabilities, pie diagrams, and odds. Both research groups concluded than these methods are seen as reliable. The same can be said of other methods like Lotteries, Betting, Certainty equivalent gambles [46]. Researchers seems to be unanimous on this point, that the use of the above tools is helpful to obtain the experts' estimations.

The main drawback of these techniques is the time taken for each probability elicited[14].

In the case of Druzdzel et al. [88] research, the average time per elicited value was around 30 min. making these techniques non viable for building large NPTs. A desirable feature of these techniques is that they do not require a knowledge of Statistics, which is often the case of the people who maintain and use the tool [71, 73, 136, 139, 151, 187].

4.4.5 Combining Qualitative and Quantitative Methods

Druzdzel and van der Gaag [61] explain a method that uses both quantitative and qualitative knowledge to build NPTs. The underlying idea is that all knowledge about the domain is valuable and can be used to constrain the model's outcome. All the available knowledge is therefore incorporated in the model in the form of a system of (in)equalities. These equations act as constraints within which lays the "true" or, in Demspter and Shafer's terms, the most "plausible" joint probability distribution.

4.4.6 Statistical Methods

Two statistical approaches can be used to build NPTs: Parametric and Non-parametric.

- Parametric approaches involve the use of distributions of known shape (for example Gaussian curves) to fit experts' knowledge (or the domain data). In general, as J. Gill [83] explains

  > ... parametric functions define what the data should look like conditioned on unknown variable values.

[14]Time, here, is interpreted in the context of eliciting large probability tables. From a different perspective, it would be difficult to assess a method in terms of time consumed given that the time taken will ultimately depend on how difficult the expert perceives the task to be [77].

The reader can use the tool called "Elicitor" to appreciate the point of this argument [126]. This tool uses graphical cues to elicit probability values: http://silmaril.math.sci.qut.edu.au/~whateley/download.html

These unknown variables are the "parameters" from which this category of approaches takes its name (in the case of a Gaussian, for instance, these would be its mean and standard deviation).

- Non-parametric approaches involve the use of distributions that have their shape determined directly by the data [36, 130] (for instance, as a frequency histogram) . These methods force fewer assumptions onto the data, but normally require more information for a reliable estimation to be obtained. They also result in more memory-intensive algorithms, because the shape of the distributions needs to be stored - this is particularly onerous in the case of high-dimensional data [196].

In this Thesis, we will focus on the parametric approaches. The main reasons for our choice are as follows:

- The availability of data in the domains of rare events in the safety critical and finance industries is more an exception than the norm.
- The complexity of modelling these domains leads to complex relationships (e.g. the causes connected to a financial fraud are multiple) which translate into large NPTs, making the use of non-parametric approaches impractical.

4.4.6.1 Eliciting Distributions' Intervals

R. Winkler [127] studied the use of Binomial distributions to elicit probability values. He focused on four different techniques to eliciting the probability value p of a Binomial distribution:

1. Cumulative Distribution (CDF) - The distribution is divided into quantiles and experts provide probability estimations for those. A Beta distribution is also used to represent expert's estimation of p; the Beta distribution is chosen from a table showing different shapes. This method is also called *quantile* method or *credible interval* method.
2. Hypothetical Future Samples - The expert begins with an assessment that it is later modified according to a change in the sample size.
3. Equivalent Prior Sample Information - Expressing prior judgments in the form of an equivalent prior sample.

4. Probability Density Function (PDF) - This technique is similar to the CDF. Both techniques require values for given intervals, only in this case it is the probability of the interval rather than the quantile. The estimation provided is relative to another interval. Experts draw a PDF according to their opinion: both assess probabilities for given quantiles.

Out of these methods the CDF and PDF provided good results eliciting probability values, with the PDF being the most successful of the two. The use of a CDF proved more challenging for the participants, who found it difficult to understand its concept. In summary, the use of quantiles and drawing the distribution were seen as successful approaches; although R. Winkler also acknowledge that the validity of these findings is limited to the use of the Bernuolli distribution in his research.

Garthwaite et al. [175] recommended two methods to elicit intervals in a PDF: "fixed intervals" and "variable intervals". The former amounts to partitioning the area into fixed intervals, e.g. 5 intervals of equal width, and eliciting values for each interval while the latter there is not a prior partition of the area. The initial interval is given by the elicited Median and subsequent interval's assessment are relative to it [172]. However, this research does not comment on how to choose the width and number of the intervals in order to produce the best results, which is something to consider given that the selection of intervals has an impact on the reliability of the results obtained [2].

Regarding the selection of intervals, O'Hagan [2] compares dividing the area into different intervals size to study how this affects the probability estimates. One approach is to divide the area into three intervals of approximately equal width, i.e: [0-33],[33-66] and [66-100]. Another approach is to divide the area into to intervals of equal width. [0-50],[50-100]. He concludes that neither approach in particular shows better results. However, he also adds that these results demonstrate the lack of research on this area.

Although eliciting intervals avoids the need to obtain directly the distribution's parameters and is more approachable for the non-statistician, it must be said that cognitive bias, like for example "over-confidence", is still present [127].

A clear example of the presence of overconfidence and anchoring biases occurs when eliciting the distribution's tails [141]. Anchoring bias results from choosing one interval, e.g. assess the Median, and make relative estimation based on that Median; if the first interval is incorrectly chosen then further estimations increase the initial error.

Tails, as we explain in Chapter 7, capture rare events; any misjudgment (e.g. overconfidence) on the tail can have an important repercussion on the forecast. For this reason, it would desirable to elicit the tails as part of a distribution and not independently as an interval. In Chapter 7 we discuss this latter point in more detail.

4.4.6.2 Eliciting Distributions' Parameters

Parametric techniques can build large probability tables eliciting only a few scenarios while maintaining the quality of the information elicited. As Druzdzel [60] comments

> ... subjects consider no more than a few scenarios, a tiny fraction of the possible scenarios in a complex problem ... The probability of any scenario within a model often can be seen as drawn from a highly skewed log-normal distribution ... What it practically means is that despite uncertainty, there is at any point usually a few very likely states of the model ... These states explain for all practical purposes almost all uncertainty.

In this section, we review the following list of distributions:

- Beta distribution
- Triangular distribution
- Pert distribution
- Binomial distribution
- Conjugate distributions
- Normal distribution
- Multivariate Normal distribution

Other techniques specific to BNs are:

- Divorcing
- Noisy-OR gates

Beta distribution. This type of distribution is often used given the different type of shapes it can produce[15]. However, it is difficult to anticipate its shape given its hyper-parameters α, β. These hyper-parameters are sometimes elicited as the probability of success (α) and the probability of failure (β) of an event. However, this interpretation does not always match the intended shape of the distribution. For this reason, a table with different Beta shapes is often used to give the expert a choice of shapes [127], see Figure 4.4.

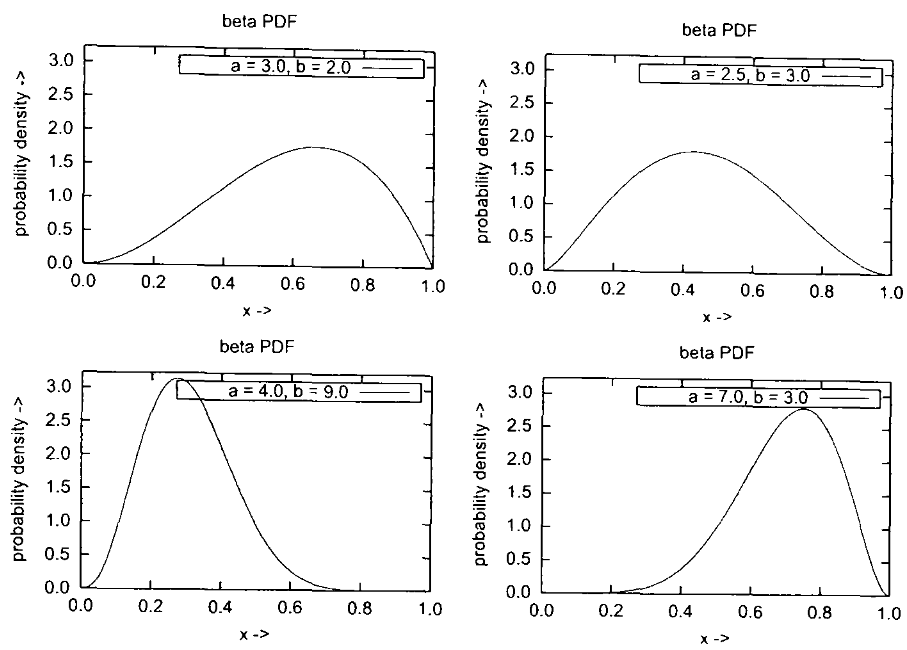


Figure 4.4: Beta distributions

Triangular distribution. As an approximation to the Beta distribution we can use a Triangular distributions, which simplifies the elicitation of the expert's knowledge. This type of compactly supported distribution is characterised by three parameters: the lower and upper bounds of the support and the mode, i.e. the point for which the probability is most likely [222] (see Figure 4.5).

This distribution does not require advanced statistical knowledge and is easy to elicit, as only three points are needed to specify the distribution. However, it constitutes a rather coarse approximation to the Beta distribution, with a tendency to overestimate the tails.

A better approximation is provided by the Pert distribution described below.

[15]Note that the Beta family can also represent a uniform distribution. Uniform distributions are used when experts do not have (or do not want to assume) prior knowledge about a variable.

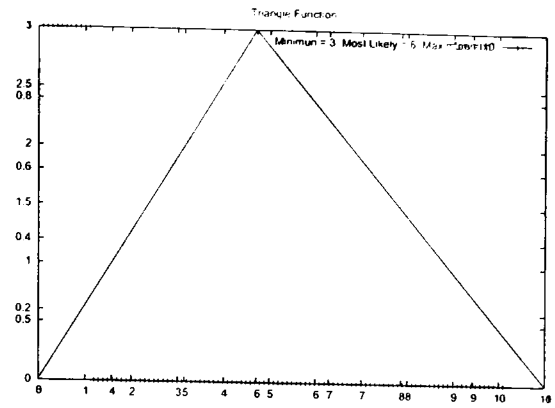


Figure 4.5: Triangular distribution. Minimum = 0, Maximum = 10, Most Likely = 5.

Pert distribution. Developed by NASA to assess time involved in developing a tool [222], the Pert distribution uses a three point estimate like the Triangular distribution but follows a smooth shape of the Beta type.

The aim is to combine the flexibility of the Beta's shape and the ease of elicitation of the Triangular distribution parameters. Figure 4.6 shows two examples of a *Pert* distribution. Here, we observe how the *most likely* outcome makes the distribution skew to either side.

Finally, we note that the Pert distribution has, like the Beta and the Triangular ones, a finite support, meaning that the probability is strictly zero outside a finite set. This assumes a knowledge that cannot always be granted when dealing with uncertainty: setting the boundaries of the support means that it is inconceivable that the event can happen beyond those bounds, thus showing a clear case of over-confidence bias (unless the finite interval is dictated by the nature of the problem). In the case of the Pert distribution, as opposed to the Triangular, the extremes of the support are not elicited explicitly, which somewhat alleviates the effects of the bias.

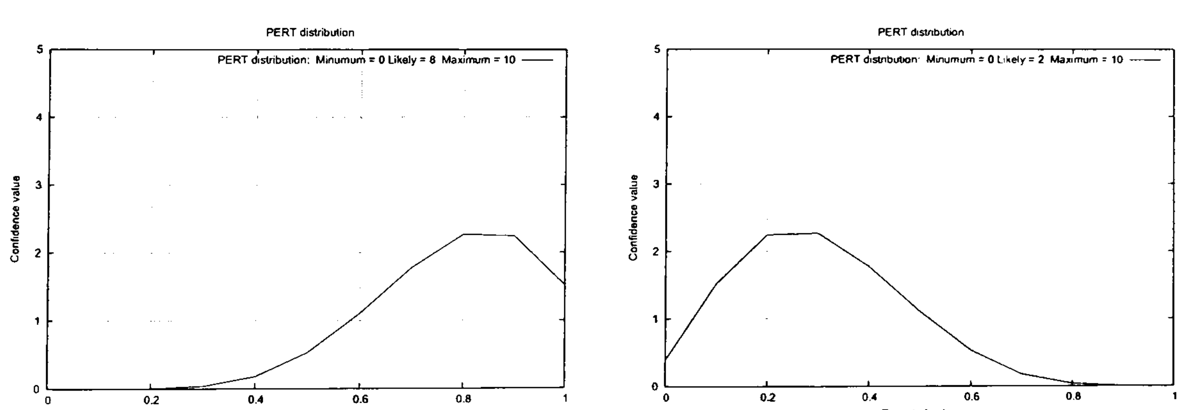


Figure 4.6: Pert distributions

Binomial distributions. P. Krause and C. Clark [124] give an example on how the Binomial distribution is applied to reduce uncertainty. In the MUNIN project, referred to earlier on page 50, they used a Normal distribution to build NPTs the mean of which is the expected value, while the deviation is given by the average of n diagnosis, each of which is distributed as a binomial with probability p:

$$\sigma = \sqrt{\frac{p(1-p)}{n}} \tag{4.1}$$

As we observe in equation 4.1 as n increases the deviation decreases, thus narrowing the central peak of the distribution; in other words, as n increases so will our knowledge about the estimation being confirmed (or otherwise, if the distribution spreads then the estimation must be modified).

P. Krause and C. Clark [124] comment

> ...once the system is in use, case data will become available. If the true state of a patient, say, is eventually ascertained, this can be used to critique the state predicted by the system.

We can see an example of this function in the OpRisk model, explained in Chapter 7, to capture the relation between the number of potential financial loss at business line level and the probability of success in reducing/avoiding them given by the control quality (e.g. quality of the internal auditing) of the business line, see Figure 4.7. In this case, the control's quality, i.e. the probability p of reducing the number of losses, is a function given by the Report Quality, the Internal and External level of auditing and the Organisational Culture that promotes risk reporting[16].

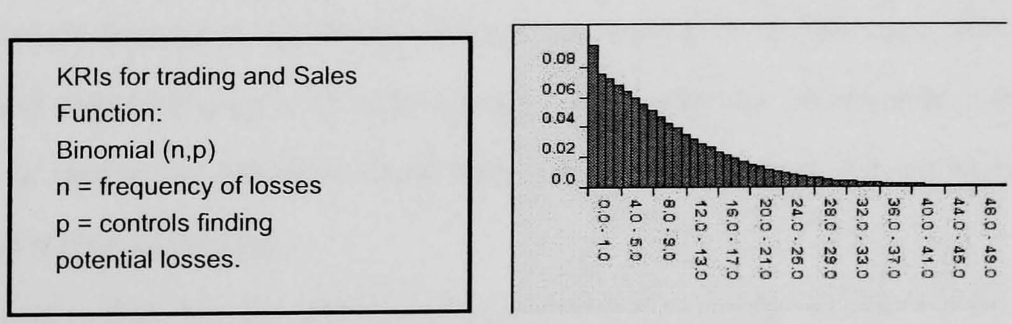


Figure 4.7: Binomial distribution for "Key Risk Indicators for Trading and Sales".

More complex parametric functions require statistical knowledge to provide the dis-

[16] See C. Alexander [7] for details on the use of this function modelling financial risk

tribution moments. In the case of Conjugate distributions, the knowledge and difficulties of eliciting different hyper-parameters for different types of distributions are reduced at the cost of assuming a prior knowledge.

Conjugate prior Sometimes it is more convenient to assume that the expert's knowledge can be captured using a specified family of distributions, this is why this choice of distributions is also called *convenience prior* [185].

Given a certain class of likelyhood functions, its conjugate prior is a class of prior distributions such that the posterior will have the same form as the prior. That is to say, both prior and posterior belong to the same distribution family [228].

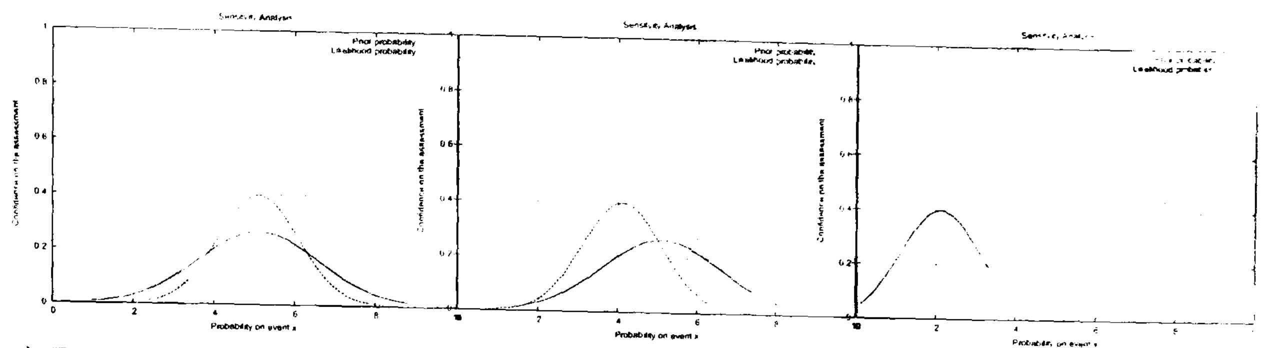
The most commonly used pair of conjugate distributions is the Beta-Binomial - the likelihood being a Binomial distribution while the prior, and consequently the posterior, are Beta distributions.

The elicitation task is reduced to choosing appropriate hyper-parameter values to capture the main features of the expert's opinion [79].

In the case of a Beta-Binomial, the experts are asked for the number of successes given a particular scenario and also for reasonable bounds on the uncertainty [172].

This prior is represented using a Beta distribution. If our prior hypothesis is confirmed by the data via the likelihood function we can conclude this relation is robust, i.e. our prior has little impact on the conditioned probability, see Figure 4.2 a). On the contrary, if prior and likelihood have widely differing shapes, this signals a problem with the elicited information, for instance due as we commented in section 4.2 to cognitive biases. In any case, this highlights a weakness of the estimation process that needs to be addressed, see Figure 4.2b). Figure 4.2c) shows as D'Agostini [49] comments, a problem with the domain being too uncertain to base any assertion: in this case, the supports of the likelihood and of the prior do not overlap. This does not necessarily imply that the assessment is wrong (as indeed it could well reflect the nature of the problem), but that it does need to be confirmed.

S. Russell and P. Norvig [197] provide a didactic example on the use of the Normal distribution in a BN model. Their example models the relationship between the *harvest*, its *subsidy* (or lack of) and the resulting *cost* of the product. The subsidy is represented by a boolean variable while the harvest is represented by a continuous variable.



a) Prior confirms likelihood. b) Prior differs from likelihood. c) Prior and likelihood contradict each other.

Table 4.2: Prior probability - likelyhood function

The parameters of the *cost* distribution are a function of the harvest; for this they use a Gaussian distribution whose mean μ varies linearly with the value of the parent *harvest* h and whose standard deviation σ is fixed.

In conclusion, two distributions are needed to model $p(c|h, \neg subsidy)$ and $p(c|h.subsidy)$:

$$p(c|h,s) = N(a_s h + b_s, \sigma_s^2)(c) = \frac{1}{\sigma_s\sqrt{2\pi}} \exp\left(-\frac{(c-(a_s h + b_s))^2}{2\sigma_s^2}\right) \tag{4.2}$$

where s indicates the presence of the subsidy, its value being either *true* or *false*. So, for $s = true$ we need to provide the parameters $\{a_t, b_t, \sigma_t\}$, and the same for $s = false$(adapted from [197]).

The interesting points we want to draw from this example are:

1. The child node becomes a function of its parent nodes: cost as a function of harvest's size and subsidy.

2. The use of a linear function to model the nature of this dependence: the mean of the distribution of the child node, the cost, becomes a linear function of the harvest: $\mu_s = a_s h + b_s$.

3. The uncertainty of the cost is captured using the standard deviation σ_s.

From these features of Russell and Norvig's example we can derive the following considerations:

1. There is no need to elicit the child node's NPT (invariably the largest table) explicitly as this becomes a function of its parents, thus making a causal explanation

easier as well. Being the distribution Normal we only need the mean and variance to produce a NPT in a short amount of time [60].

2. The assumption of linearity is pertinent as long as the variables remain monotonically related, and there are no abrupt changes or discontinuity in their functional dependence: in this case, a simple linear model will provide good results [22]. Serve as an example G. Wright and P. Ayton's [76] comments on a study on Hodgkin's disease in which the doctors coded the patients' biopsies and made an overall rating of severity:

   > These overall ratings were very poor predictors of survival time. But the variables provided by the doctors were used in a standard linear regression analysis. These made excellent predictions.

   In fact, even if the underlying relationship is not linear, making this approximation (unless the nonlinear dependence is crucial to the problem at hand and cannot be ignored) produces better results than using more complex approaches[17].

3. The assumption of variable independence in the linear model is also consistent with BN assumption of independence.

Notice that the parameters a,b and σ in Russel and Norvig's example are estimated from the data and that σ is fixed.

In the ranked node approach we assume the absence of data and instead we use experts' knowledge. The values a,b are elicited Following this approach This can be done by assuming that Y is normally distributed around the value provided by the regression model, with a given standard deviation σ:

$$P(Y = y) = Normal(y|\mu,\sigma) \tag{4.3}$$

where $\mu = \beta_x x + \beta_z z$. This is a valid assumption as long as the noise on the estimate

[17] There is an interesting philosophical discussion regarding complexity vs simplicity, which supports the case for simple models [146] (in the author's view, many of the arguments could be extended from the topology of the model to elicitation methods). In support to this thesis, we quote *Darwin's Theory of Evolution*: a model must be general enough to adapt to changes on the domain, thus, simpler models have good chances to fit future demands. Another relevant principle is *Ockham's razor*: all other things being equal, choose the simplest explanation. More complex explanations are often a sign that a problem is not well understood.

of Y can be interpreted as the sum of independent and identically distributed *expert uncertain assessments.*

It is worth noting that when the distribution to be elicited is approximately symmetric experts' estimates have shown a high degree of accuracy [79]: this suggests that the experts are better at providing probabilities when these are in the shape of a Normal distribution [172].

On the other hand, if the resulting distribution is not symmetric experts' estimates are poorer. However, as Garthwaite's [175] research has shown, there seldom is enough difference from the point of view of accuracy to justify assuming an asymmetric distribution for the expert' opinion.

In the case of eliciting the standard deviation, research shows that the experts give poor estimations [175]. For this reason we need alternative ways to obtain this quantity:

- using a verbal/numerical table (see section 4.4.2);
- indirectly by analysing the factors that contribute to that variance;
- using statistical distributions (e.g. see section 4.4.6.2);

In the indirect approach the standard deviation is a function of the parent nodes, these represent the systematic errors and random errors that can derived from the elicitation exercise [3], e.g. variance = systematic errors + random errors. In this way the standard deviation can be dynamically adjusted depending on the novelty of the domain, the experts' knowledge of the subject and possible bias, among others. Figure 4.8 illustrates this idea.

Multivariate Normal Another particular case, studied by Garthwaite et al. [14] and Kadane et al. [183], is the use of a Multivariate Normal distribution to represent more complex relationships involving more than one variable. (In Chapter 8 section 8.3 we show Multivariate method consistent with the node approach.)

However, our view is that, if at all possible, one should re-cast the relationship in a way that avoids the multivariate dependence, in order to make it more amenable to elicitation.

We agree with O'Hagan [156] when he comments

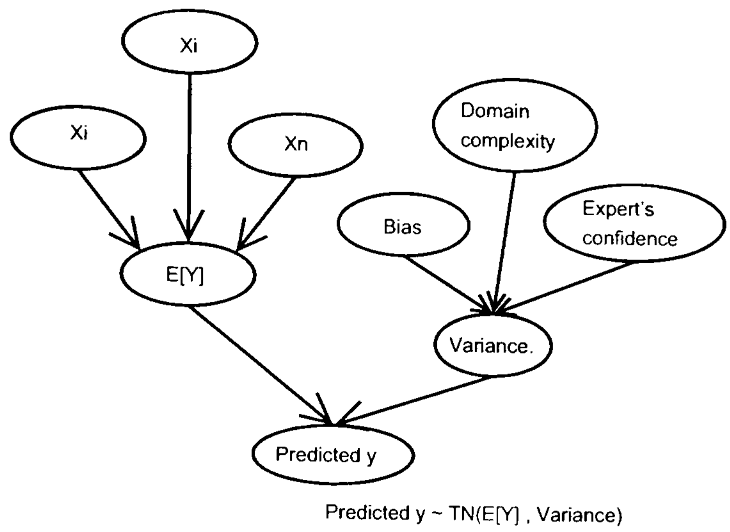


Figure 4.8: Mean and variance as a function respectively of the data and of the explicitly defined sources of uncertainty.

> ... Even with two variables, these are more complex things for the expert to think about, and with three or more variables it will be very difficult for the expert to assess suitable values for joint or conditional probabilities ...

It is generally possible to re-interpret the model in such a way that node's relationship becomes univariate. As argued by Kadane and Wolfson [117],

> If complex probability is required decompose the problem into eliciting simpler probabilities and combine then using probability laws, using influence diagrams they can appreciate more the value of their input ...

The next section explains how to decompose a problem domain using a technique called *divorcing.*

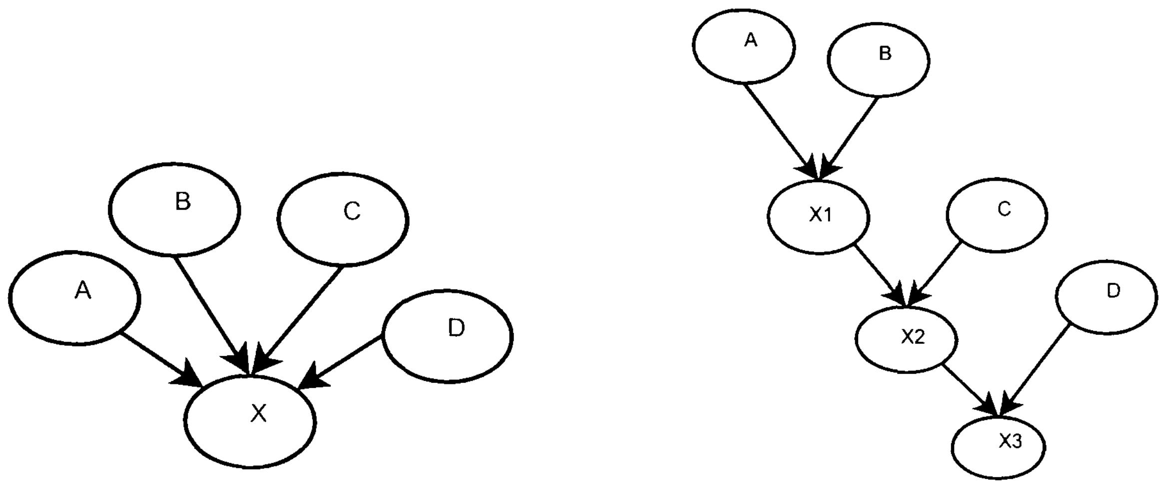
4.5 BN techniques to reduce the size of the NPT

There are essentially two specific approaches to reducing the size of the NPT:

1. Changing the graphical structure of the BN (namely, *divorcing*) or
2. *noisy-OR* gates.

4.5.1 Divorcing

The first approach builds on the principle of *divorcing* parents by introducing intermediate variables, thus reducing the size of the node probability Table [111]. This approach can also be used as a tool to facilitate and improve expert elicitation. As we commented earlier in section 4.4.6, experts are poor at estimating the compound effect of two or more variables [156]. Garthwaite et al.'s [79] research shows that experts tend to underestimate the effect of single variables showing overconfidence when estimating their combined effect: under this respect divorcing can be seen as a modelling tool.



(a) Node X with four parents $pa(X) = \{A,B,C,D\}$. Assuming they are all binary variables, we have $2^{4+1} = 32$ probability values to elicit for the node X .

(b) Same network as in Figure (a) but this time we have *divorced* the parent nodes taking them one at a time. In this way, we only have $2^{2+1} = 8$ probability values to elicit per intermediary variable X_i where $i = 1,2,3$. Thus, the total values elicited are $8 \times 3 = 24$ values.

Figure 4.9: Divorcing is a techniques to reduce the size of an NPT by introducing intermediate variables.

Figure 4.9 shows an example of divorcing. Note that the way the parents nodes are grouped in Figure 4.9(b) is irrelevant to the final computation[18]. That is, we might have chosen a different order: $\{B,\ D\}$, $\{A\}$ and $\{C\}$; as long as the order respects the (in)dependency among nodes.

In any case, the way in which the nodes are grouped should facilitate their elicitation. The intermediary variables X_1,X_2 and X_3 appear only for computational purposes, and have no other (causal) meaning attached to them. In terms of elicitation, they represent the compound effect of their parents.

[18]There may be some disparity between results but this is due to the computational algorithm used

4.5.2 Noisy-OR Gates

The second approach, called *noisy-OR*, was introduced by Pearl in [178]. This approach models the case where the relationship between parent and child is uncertain or as Pearl would say, the causal relationship between parent and child may be *inhibited*.

We will introduce this technique by means of an example in a medical context (where this approach has mainly been applied[19]), before giving its mathematical definition.

Let us assume that we want to discover the cause of a patient's fever. Suppose that there are only three possible causes, i.e. *cold*, *flu*, and *malaria*. However, we can think of a patient who has a cold but who does not exhibit a fever, in which case the causal relationship would be inhibited. Of course, the conditions that inhibit cold as a cause are different from those that inhibit flu, and so forth. Also, we assume that if the patient shows a fever due to a cold then the *inhibitors* associated with cold must be inactive, and that the same applies to the other causes. With these assumptions we only need to elicit three probability values to build the *noisy-OR* table 4.3, namely

$P(\neg fever|cold, \neg flu, \neg malaria) = 0.6$

$P(\neg fever|\neg cold, flu, \neg malaria) = 0.2$

$P(\neg fever|\neg cold, \neg flu, malaria) = 0.1$

(the numerical values are those used in Table 4.3 below).

Cold	Flu	Malaria	P(fever)	P(¬fever)
f	f	f	0	1
f	f	t	0.9	0.1
f	t	f	0.8	0.2
f	t	t	0.98	**0.2*0.1=0.02**
t	f	f	0.4	0.6
t	f	t	0.94	0.1*0.6=0.06
t	t	f	0.88	0.6*0.2=0.12
t	t	t	0.988	0.6*0.2*0.1=0.012

Table 4.3: Example of noisy-OR, from [197]

Pearl [178] defines *noisy OR* as follows. Let $\mathbf{U} = \{U_1, U_2, ..U_n\}$ be the n parents of node X and $I_1, I_2, ..I_n$ be the associated inhibitors of these parents nodes, respectively. Let q_k stand for the probability that the *k-th* inhibitor is active. If U_i is the only parent that

[19] See for instance Agnieszka [169] for an example of the use of Noisy OR-gates in the diagnosis of liver disorders.

is in the "true" state, X will be "true" if and only if the inhibitor associated with U_i (i.e. I_i) remains inactive. That is,

$$p(X = true|\ U_i = true,\ U_k = false \forall k \neq i) = 1 - q_i$$

As we have seen in our example, the Noisy-Or operation makes two assumptions [178]:

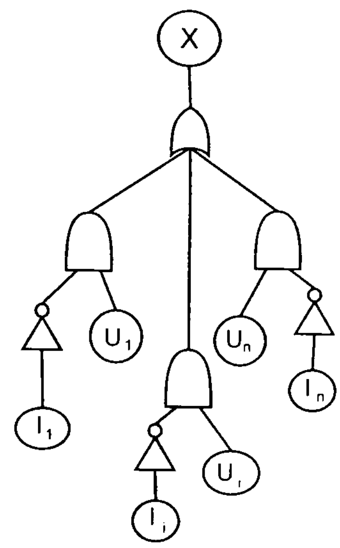


Figure 4.10: Logic diagram showing the concept of a noisy-OR gate. Adapted from [178]

1. Accountability: An event E (e.g. having a fever) is presumed false if all the conditions listed as causes (e.g. malaria, flu, cold) of E are false. This of course implies that all the causes must be listed[20].

2. Independence: the inhibition of each parent is independent of the inhibition of the other parents (e.g. the inhibitor of cold is completely unrelated to that of malaria). Therefore,

$$\begin{cases} p(X = false|U) = \prod_{i|U_i = t} q_i \\ and \\ p(X = true|U) = 1 - \prod_{i|U_i = t} q_i \end{cases}$$

where $\{i|U_i = t\}$ indexes the subset of parents that are true.

For instance, in table 4.3 we observe that Fever is false if and only if all of its parents in the "true" state are inhibited.

[20] An interesting extension of this concept is the *leaky Noisy-OR*. This is for the case in which the effect can still materialize even if none of the parents are active. The leaky node stands for the *other causes* not captured in the model.

Because of the independence assumption, the probability of this happening is the product of the probabilities of each parent being *inhibited*:

for example, $p(Fever = false|Cold = false, Flu = true, Malaria = true) = 0.2 * 0.1$.

Using *noisy-OR* we only need to define n probability values (where n stands for number of parents), instead of the 2^{n+1} values that would be needed otherwise. Figure 4.10 illustrates this concept using a logic graph.

Notice that in the case of *noisy-OR* variables must be Boolean, i.e. the causes must be present or absent. A generalisation called *noisy-AND* can be used when causes are not only present or absent but they can also be present with different degrees of intensity, e.g. low, medium, high [55].

We can apply this technique in the field of risk and safety assessment whenever the assumption about the inhibitors acting independently is reasonable, e.g. an air-traffic controller's advice to avoid a mid-air collision and pilot's skills to perform the required manoeuvre. In this case, both inhibitors are independent. An altogether different scenario would be an airplane radar system and the radio equipment failing due to a power failure. In this case, both events have a common inhibitor, thus the *noisy-OR* model cannot be applied.

4.6 Discussion

In this section, we briefly discuss the above elicitation and modelling techniques in the light of the following key points:

- the expert's judgment is a genuine source of information and should be taken into account in the absence of data or in order to complement the data;
- the expert's time is scarce and costly;
- the expert's knowledge of the domain is limited[21];
- the users more often than not have little statistical knowledge;
- a simple approach makes the user understand and follow a model's forecast in an informed way, as opposed to seeing it as an Oracle [60];

[21] O'Hagan [155] questions whether eliciting more summaries from experts would result in a more accurate, true distribution. In his opinion the experts' accuracy is limited; eliciting more summaries over a variety of distributions would not improve the results.

- designing a BN is an iterative process through which we add, remove or modify parts of the topology and/or the probability estimations as the modelling progresses.

We have discussed earlier that subjective judgment is a genuine source of information, potentially as valid as an objective judgment and even more so when the latter is absent: trying to use other approaches like Objective Bayesian in an attempt to grant validity to the model is what O'Hagan called a *heresy* [155].

In the case of BNs, probability theory grants consistency to the elicited values. Calibration is ascertained as feedback and data become available, and with the help of techniques such as sensitivity analysis.

Also, expert elicitation is facilitated by the concept of conditional dependency, that indicates which are the relevant nodes to elicit and supports the interpretation of a BN as a causal network.

It will be interesting to see how the development of ontologies will affect the discussion on subjective elicitation. We can see the use of ontologies in the context of defining network semantics and consolidating model findings. During our project with the air-traffic experts, explained in Chapter 6, we ran several times into misunderstandings about node definitions; i.e. expressions that are semantically identical but syntactically different or syntactically similar but semantically different. For instance, there was a misunderstanding caused by the use of the word "pilot" in one of the model's nodes. For us this term represented the pilot's performance, while to the air traffic controllers "pilot" meant the performance of the plane's crew; consequently, probability estimations had to be changed. Using ontologies we can avoid this type of problems by explicitly defining each node's meaning; definition that is shared and accepted by the experts. Such definitions can greatly reduce the ambiguity on what is being said.

However, as Gruber [85] pointed out, the level of abstraction in a definition (and hence its level of ambiguity) is given, to an extent, by the domain it captures. If in the domain of engineering concepts are formulated as mathematical expressions learnt by the engineering community, can we say the same about the concepts related to financial risk?

Regarding probability elicitation methods, we cannot see the viability of graphical tools such as the probability wheel or histograms to build large NPTs, mainly due to the time required for eliciting the probability values in this way. Such tools can be used advantageously in the context of eliciting specific priors. e.g. in case of disagreement between two experts on a distribution that needs detailed attention. a situation that suggests that the underlying relationship could be bi-modal.

As part of this thesis, we implemented a Bezier Curve and a Histogram as eliciting tools, see Appendix A. This type of curve was already used by the RADAR group during the Serene project [37]. Unfortunately. the author cannot report on how useful. or otherwise, both tools are given the lack of opportunities we had to test them.

Verbal scales deserve a special mention. We can find authors that recommend their use. while others criticise their shortcomings. On one hand we have the case of eliciting single values using an intuitive method, but on the other hand we have the ambiguity of the semantics, the potential bias of how an expert can interpret a scale. the lack of consistency between scales, and the ad hoc mapping of probability. Nevertheless. we find researchers that still use them: Bolger [22] recommends their use in particular cases. Budescu [216] recognises them as a valid tool that experts can choose to communicate their opinions, van der Gaag et al. [217] finds them a useful tool in the domain of medicine and to elicit many probabilities [219]. Renooij and Witteman [193] report on how the use of a numerical scale coupled with a linguistic scale could improve expert assessments.

From the RADAR group's experience, we view this tool as a viable option to elicit probabilities which are later confirmed and/or refined as data is accumulated.

Parametric distributions can produce NPTs in a short amount of time. It is often the case that there is more than one potential distribution family to capture a NPT. The type of distribution used, e.g. Beta, Binomial, Normal is left to the expert/analyst's choice and this is determined by whichever can provide the *best fit*. The selection is driven by the expert's knowledge in Statistics, whether the elicitated estimations are observable quantities [117] or repeteable events [22], and by the care to avoid possible bias that can result from overly complex distributions (for example, in the Multivariate Normal distribution the compound effect of different variables is affected by over-confidence, underestimating the contribution of each individual variable [79]). The choice is of course constrained also by practical considerations, such as the time and budget available.

Given that the distribution chosen is conditioned on such mundane factors such as amount of time available, costs, and limited expertise among others, to what extent does the model benefit from dedicating more time (and resources) to the elicitation of more demanding statistical distributions? On this issue, Garthwaite's [79] research concluded, as we have seen, that in the case of eliciting skewed distributions the benefits obtained did not outweight the additional effort required by modelling such distributions.

Henrion et al.'s [99] research answers this question from a different perspective by analysing to what extent the sources of imprecision in the quantification of each node's relationship have an impact on diagnosis networks:

> The addition of massive amounts of random noise to the link, leak, and prior probabilities produced only modest decrements in diagnostic performance ... It is more important to identify findings, and diseases ... and their relationship, than to quantify the relations with high level of precision.

These findings are also corroborated by the work of Chan and Darwiche [29]. It is also true that these conclusions depend on the type of domain and node relationship [155]; some relations can prove more *robust* than others.

It is therefore for these reasons that we advocated a general, flexible approach that can accommodate different types of relationships, rather than a bespoke distribution whose development would be lengthy and costly. We support a simple approach that would captures the experts' opinions, and not their knowledge in Statistics.

Although such a general approach may be seen as coarse, it originated from the extensive experience of the RADAR group in the development of real world BNs [71, 73, 136, 139, 151, 187].

The next chapter explains this approach in detail.

Chapter 5

Building Node Probability Tables

5.1 Introduction

This Chapter benefits from the work and experience of the group RADAR, lead by N. Fenton and M. Neil, has producing solutions to a number of real-world problems such as

- safety of embedded systems in the railway industry [187],
- military vehicle reliability [139] and
- software defect prediction in consumer electronics products [71, 73, 136, 151].

These applications involved building large-scale BN models. As a result of the difficulties that we encountered in BN model building, we were well aware of the limitations of relying on purely "hand crafted" approaches, like elicitating NPTs using a table of different Beta distributions, in which each variable and each NPT had to be elicited exhaustively with domain experts; especially for nodes with many states which are convenient during elicitation as Andreassen et al.[1] [12] points out

> .. nodes with multiple states has given a conceptual simplicity that makes knowledge acquisition and verification easier.

This Chapter focuses on only one especially important part of this problem: how large NPTs can be built for a commonly occurring class of nodes called "ranked" nodes

[1] Note that in the case of the MUNIN project nodes have up to nine states [12].

(which represent qualitative variables that are mapped to intervals in a continuous distribution).

We begin in section 5.2 by outlining the problem. In section 5.2.2. we formalize the notion of ranked nodes, along with the conditions under which they occur most commonly in BNs. Section 5.3 explains the previous step to using a Normal distribution to express expert's uncertainty. Using what we call the *odds* function we were able to calculate the deviations of the expert's estimates. In sections 5.4 and 5.5, we describe the class of causal weighting functions required to generate the NPTs for these ranked nodes. Our method is based on the representation of NPTs by means of parametric probability functions, where the child node's probability is defined as a weighted function of the parent node values. The weighted rank node functions specified herein (which turn out to be sufficient for most applications) are:

- Mean Average,
- Minimum,
- Maximum

In section 5.6, we describe the other instance where ranked nodes commonly occur, namely, as indicator nodes.

The discussion on this Chapter is postponed to Chapter 8, after introducing the two examples used to demonstrate the validity of this approach.

5.2 The problem and background

5.2.1 Ranked nodes

Consider the BN fragment shown in Figure 5.1. Such fragments are very typical of those that frequently occur in the real world models already cited.

They are characterized by the fact that node values are typically measurable only on a subjective scale like {very low, low, medium, high, very high}, and only extremely limited statistical data (if any) is available to inform the probabilistic relationship for Y, given X1 and X2. However, there are significant expert subjective judgements that can be used.

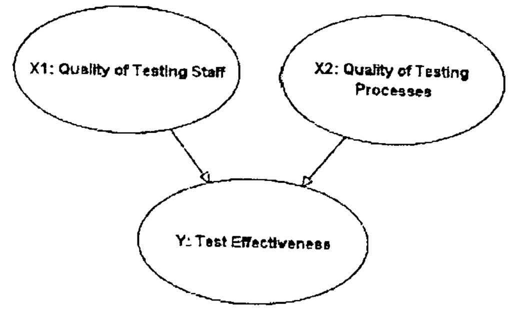


Figure 5.1: A typical BN fragment

Assuming that each of the nodes has five states the NPT for the node Y has 125 states. This is not an impossible number to elicit exhaustively, but from extensive experience, we know that all kinds of inconsistencies and arise when experts attempt to do so. If the number of states increases, and/or the node Y has additional parents, then exhaustive elicitation becomes infeasible, especially as real-world models invariably involve dozens of fragments like these.

Hence, the problem and challenge is to produce an appropriate NPT for the node Y that makes the most of limited expert elicitation unless we adopt Wellman [224] approach and avoid the problem of eliciting NPTs altogether. This type of network uses *qualitative influences* that is, assigning signs {+, - , ?, 0} to the node's links instead of assigning probability estimations.

5.2.2 The nature of Ranked nodes

Ranked nodes represent discrete variables whose states are expressed on an ordinal scale that can be mapped onto a bounded numerical scale that is continuous and monotonically ordered. We can assume that all ranked nodes are defined on an underlying unit interval [0-1] scale. For a given number of intervals defined and labeled, on this scale, we simply discretized accordingly, see Figure 5.2.

As far as the user is concerned, the underlying numeric scale is invisible, the displayed scale is still the labelled one rather than the numeric one, but the latter is used for the purposes of computation and generating the NPT. Note that BN software tools like Hugin [13] do not provide this type of functionality. For this reason we needed to implement a computer program to make the BN construction and editing task much simpler than is otherwise possible. In particular, provided that they appear in the appropriate combinations described below, the normally complex task of constructing sensible

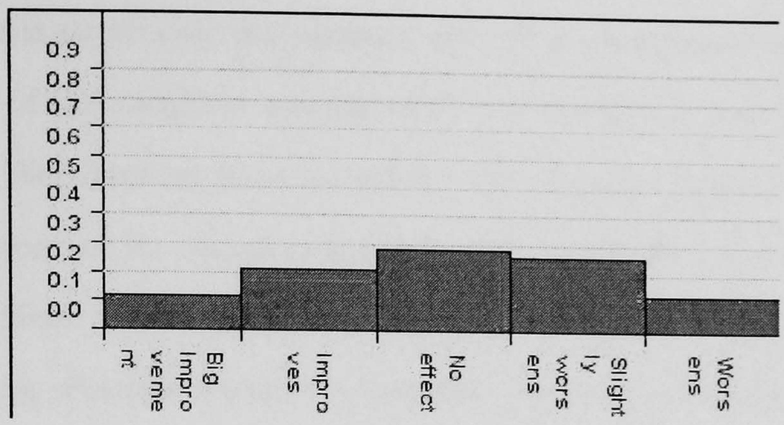


Figure 5.2: Ranked node. The probability range [0-1] is divided into five intervals and labeled as *Big Improvement, Improves, No effect, Slight worsens, Worsens* each interval representing 20% of the distribution. Thus "Big Improvement" is associated with the interval [0 - 0.2), "Improves" is associated with the interval [0.2 - 0.4) and so forth.

associated NPTs is drastically simplified.

From the experience of our previous applications, experts typically wanted to complete an NPT by using a simple averaging scheme to compute the maximum or minimum value as a guide to defining the "central tendency" of the child node based on a set of causal parent node values. For instance, during the NATS model (see Chapter 6 section 6.7 page 111) we adopted an approach based on sampling values to construct the NPT for a node like Y which resulted in expert elicitation assertions like the following:

- When X_1 and X_2 are both "very high", the distribution of Y is heavily skewed toward "very high".
- When X_1 and X_2 are both "very low", the distribution of Y is heavily skewed toward "very low".
- When X_1 is "very low", and X_2 is "very high", the distribution of Y is centered below "medium".
- When X_1 is "very high", and X_2 is "very low", the distribution of Y is centered above "medium".

Since we are assuming that each node has an underlying numerical scale in the interval [0, 1], such assertions suggest intuitively that Y is some kind of a weighted average function. In fact, experts found it easier to understand and express relationships in such terms. Many so-called "self-assessment" or "scorecard" [21] systems are based around little more than the weighted averages of attribute hierarchies.

The challenge is to provide the appropriate BN implementation that captures the explicit simplicity of the weighted average while also preserving the intuitive properties that the resulting distributions have to satisfy. For example, simply making Y the "exact" weighted average of its parents does not work, since the only uncertainty in the distribution of Y, given its parents, is the result of discretization inaccuracy rather than deliberate modelling. For this reason we needed a function to capture the uncertainty of this relationship. The next section explains our first attempt to explicitly model uncertainty.

5.3 Odds function

The odds function is defined in equation 5.1:

$$Y = \frac{1}{odds^{|Y_{[E]} - Y_i|}} \tag{5.1}$$

where Y is a ranked node and $Y_{[E]}$ stands for the expected value for Y and Y_i for the labeled states of Y, for instance $Y_{very\ low}, Y_{low}, Y_{medium}, Y_{high}, Y_{very\ high}$ and corresponding ordinal values $\{1, 2, 3, 4, 5\}$ respectively. The term *odds* stands for the expert's confidence in his assessment.

From equation 5.1 we observe that expert's uncertainty is inversely proportional to the *odds* raised to the power of the "distance" from the expected value $Y_{[E]}$ and the rest of possible states Y_i. As expected, as we move away from the expected value $Y_{[E]}$ the probability decreases.

To obtain the *odds* value we provide a look up table where the expert can map his confidence using a verbal/numerical scale as in Table 5.1. A lower *odds* value produces a peaky distribution shape meaning a stronger confidence in the estimate while a high value would have the opposite effect; a wider spread distribution.

For instance, using the BN fragment on Figure 5.1. Let us assume that the expert is "confident" that when the staff and testing quality are both "very low" the test effectiveness is also "very low". In our case, this "confidence" translates as *odds* = 2. Using equation 5.1 we obtain $Y = \frac{1}{2^{|1-1|}} = 1$ for Y being "very low", see Table 5.2. The last step is to normalise the values obtain on this table, see the right most column of Table 5.2.

This process was easy to compute and to elicit. However, after extensive testing, N.

Confidence	*odds*
Almost Certain	1.1
very confident	1.5
confident	2
likely	2.5
not sure	3
uncertain	4

Table 5.1: *Odds* look up table.

Y	X1 and X2 = "very low"	Normalise
very low	1	0.5
low	0.5	0.25
medium	0.25	0.13
high	0.125	0.064
very high	0.03	0.015

Table 5.2: Estimates for child node using Odds function

Fenton realised that the symmetrical property of this function did not hold in the case of inter-causal reasoning.

For example, suppose we have observed Y and $X1$ and wish to *explain away* the value of $X2$ as follows: If Y is "very high", and $X1$ is "very low", then we would be almost certain that $X2$ is "very high". If Y is "very high", and $X1$ is "average", then we would be confident that $X2$ is "very high" but not as confident as in the above case. However, the odds function failed to satisfy this property.

A straightforward solution that fulfills this property was the use of the Normal distribution. The challenge here, however, was to implement this function in a computer program that was the reason of using the odds function in the first place because its simplicity to implement. We were able to make such an implementation (see Appendix); now part of the Agena Risk tool [134]. The Normal distribution was later dropped in favour of a bounded Normal distribution which is explained in the next section.

5.4 Modelling Ranked Causes Using a Doubly Truncated Normal Distribution

Formally, the ranked node's causal structure is characterized by a joint probability distribution with a set of causes X containing $i = 1, 2, .., n$ ranked nodes $\underline{X}$ as parents of Y

:

$$p(\underline{X},Y) = p(Y|X)\prod_{i=1}^{n} p(x_i)$$

In general, the node Y is considered to be a *consequence* of two or more *cause* nodes, where each of the cause nodes is assumed to be independent when calculating the NPT. The BN in Figure 5.1 is a very simple example.

We can draw an analogy with linear regression, where $Y = X1\beta + X2\beta + \varepsilon$, if we consider the child node Y the result of a linear relationship between the parent nodes, with ε approximating a Normal distribution of mean 0 and variance σ_y^2 (written $N(0, \sigma_y^2)$, and where the contribution to the variance of Y is σ_y^2).

The regression analogy is apt, since we are attempting to "target" the area of central tendency in Y, given different values of X, and then are adding a fixed amount of uncertainty around this. The only issue that we need to resolve is the contribution of each cause to the effect, and a clear way to do this is to use the correlation between the cause and the effect as the appropriate measure.

Rather than the Normal distribution commonly assumed in linear regression for ranked causal nodes, we use the doubly truncated Normal distribution (denoted TNormal hereafter) as defined, for example, in [42], where all nodes are truncated in the [0,1] region.

Unlike the regular Normal distribution (which ranges from $-\infty$ to $+\infty$), the TNormal has finite end points. We denote the TNormal by $TNormal(\mu, \sigma^2, 0, 1)$ where μ is the mean, and σ^2 is the variance.

In the TNormal, we start with a regular Normal distribution but "ignore" the probability mass to the left and right of the finite end points and then normalize the resulting distribution over the finite range [0, 1]. This enables us to model a variety of shapes, including a uniform distribution, achieved when the variance $\sigma^2 \to \infty$, and highly skewed distributions, achieved when $\sigma^2 \to 0$.

We use a simple weighted sum model to measure the contribution of each X_i to explaining Y as a "credibility weight" W_i (it can also be elicited from an expert in this way) expressed as real values $W_i \geq 0$. The higher the credibility index, the greater the correlation between X and Y . Thus, in our method, the equivalent to the error variance σ_Y^2 in the linear regression model is simply the inverse of the sum of the weights:

$$\sigma_Y^2 = \frac{1}{\sum_{i=1}^{n} W_i} \quad for \ \sum_{i=1}^{n} W_i > 0$$

Given that Y lies within [0, 1], we must normalize the regression equation $E(Y) = \sum_{i=1}^{n} XiWi$ by dividing by $\sum_{i=1}^{n} Wi$. Thus,

$$p(Y|\underline{X}) = TNormal \left[\frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i}, \frac{1}{\sum_{i=1}^{n} W_i} \right] \tag{5.2}$$

Suppose, for example, that $n = 3$ and that the allocation of weights W_i for each Xi's contribution to explaining Y is in the ratio 2:3:5, with a resulting variance $\sigma_Y^2 = 0.1$. Then, the joint distribution generated will be

$p(Y|\underline{X}) = TNormal \left[\frac{2X1+3X2+5X3}{10}, 0.1, 0, 1\right]$

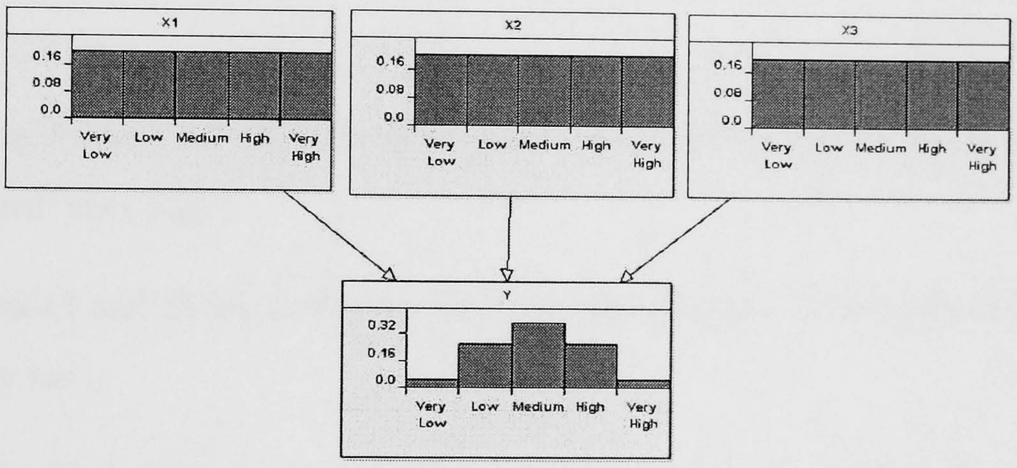


Figure 5.3: WMEAN function for Y, given $X1$,$X2$ and $X3$.

The resulting distribution and BN model are shown in Figure 5.4. The resulting distribution for $p(Y)$ will not produce summary statistics exactly matching the function because we are using coarse discretization in arriving at the result.

Given this, the mean values will tend to differ within the interval range specified. Specifically, for five ranks defined on $[0-1]$, the mean value may be out by up to 0.1. Also, the variance values observed will be considerably higher because of the coarse discretization. However, neither of these are major problems, since the aim is to produce a good fit to the expert's distribution rather than a good approximation to a TNormal distribution.

5.5 Modelling Ranked Causes Using Weighted Min and Max

Sometimes the relationship between parent and child nodes is one where one parent's input "overrides" the impact of the other parent(s). We are going to explain this point using an example taken from our project with NATS explained in Chapter 6.

In Figure 5.4 we observe that the child node "TCAS & Pilot's performance"(Y) depends on "Pilot's skill" ($X1$) and "TCAS advise" ($X2$). In this case, we elicited the following information:

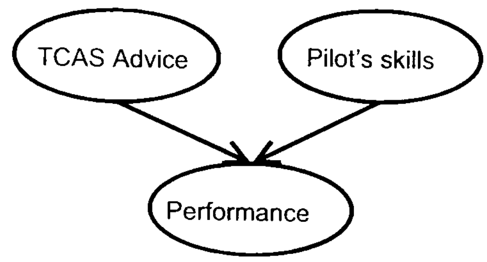


Figure 5.4: The node "TCAS and Pilot's performance" depends on two causes: "TCAS advise" and "Pilot's Skill".

- When $X1$ and $X2$ are both "very high", the distribution of Y is heavily skewed toward "very high".
- When $X1$ and $X2$ are both "very low", the distribution of Y is heavily skewed toward "very low".
- When $X1$ is "very low", and $X2$ is "very high", the distribution of Y is centered toward "low".
- When $X1$ is "very high", and $X2$ is "very low", the distribution of Y is skewed toward "low".

Figure 5.5 illustrates the result of eliciting, with the aid of graphs, a particular set of sampled scenarios for the child node. On scenario "A" we observe that when TCAS advised is considered to be "Very High" (understood in this case as accurate) and Pilots Skills are "Very High" (understood as excellent skills) Performance is expected also to be "Very High".

To make the analysis of this drawings easier we are going to use the table 5.3. There, we observe that whenever TCAS or Pilot's skill are expected to be both "Very Low" child

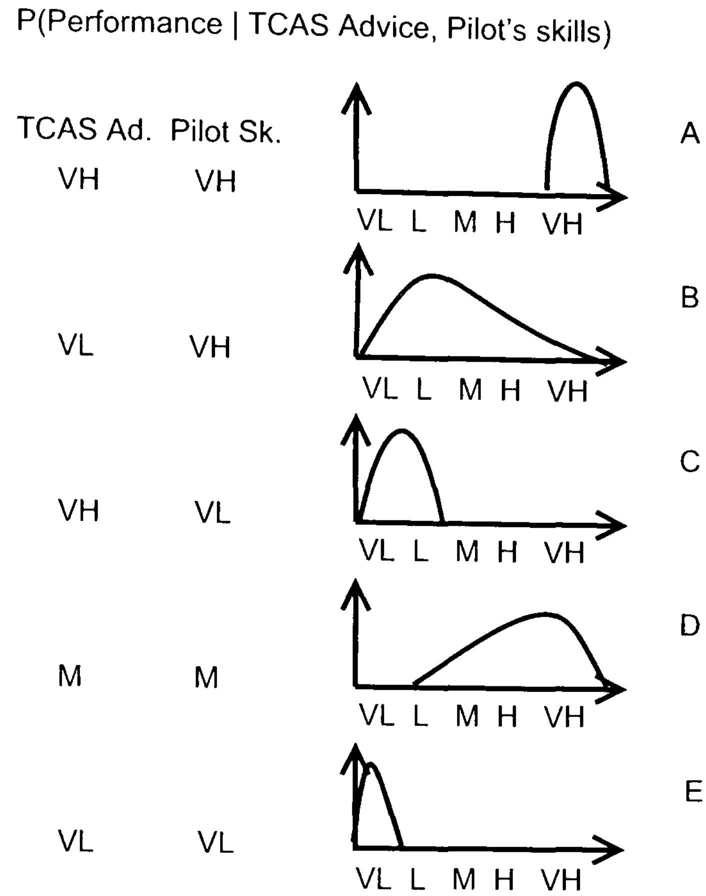


Figure 5.5: Expert's drawings of a sampled set of scenarios.

TCAS advice	VH	VH	VL	VL
Pilot's skill	VH	VL	VH	VL
Performance				
VH	VH	-	-	-
VL	-	VL	VL	VL

Table 5.3: Relationship table between TCAS, Pilot and Performance

node Performance is also expected to be "Very Low", that is, the expected estimation is skewed toward whichever parent has the *lowest* state.

In this case, the experts seem to be more concern with the situation where one of the parent nodes has a high probability of not performing well (*regardless* how well the other parent performs), e.g. TCAS gives wrong advise or pilot's skills being poor.

We are aware that this may result in a too conservative estimate but given the nature of the domain, mid-air-collision, a cautionary approach is justifiable.

So, a weighted sum for the child node will *not* produce a NPT to satisfy these elicited requirements. Formally, child distribution's mean is something like the *minimum* of the parent values, but with weighting in favour of $X1$. Comparing scenario "B" and "C" on Figure 5.5 we can see that "Pilot's skill" has more weight on the resulting distribution shape. This makes sense if one thinks that it is ultimately pilot's responsibility to fly the plane and to follow any advice. The necessary function, which we call the *Weighted Min function* (WMIN), has the following general form,

$$WMIN = MIN_{i=1,2...n}\left[\frac{W_iX_i + \sum_{i\neq j}^{n} X_j}{W_i + (n-1)}\right] \tag{5.3}$$

where the weight $Wi \geq 0$ and n are the number of parent nodes, with a suitable variance σ_Y^2 that quantifies our uncertainty about the result, thus giving

$$p(Y|\underline{X}) = TNormal\left[WMIN(X), \sigma^2, 0, 1\right] \tag{5.4}$$

The WMIN function can be viewed as a generalized version of the normal *MIN* function. In fact, if all of the weights Wi are large, then $WMIN$ is close to MIN. At the other extreme, if all the weights $Wi = 1$, then $WMIN$ is simply the average of the $X$is. Mixing the magnitude of the weights gives a result between a MIN and an $AVERAGE$. In the above example, taking $W1 = 3$ and $W2 = 1$ (with a variance $\sigma_Y^2 = 0.01$) yields the results, as shown in Figures 5.5 and 5.5.

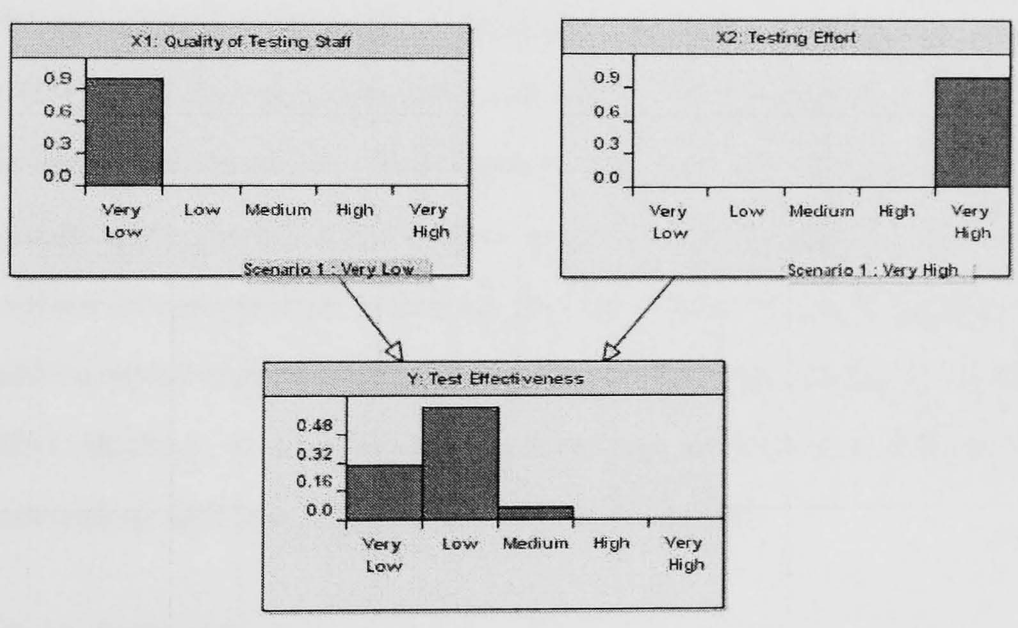


Figure 5.6: WMIN function for Y. "TCAS advise" = "very low", with $W1 = 3$. "Pilot's Skill" = "very high", with $W2 = 1$

We can also use an analogous $WMAX$ function see equation 5.5:

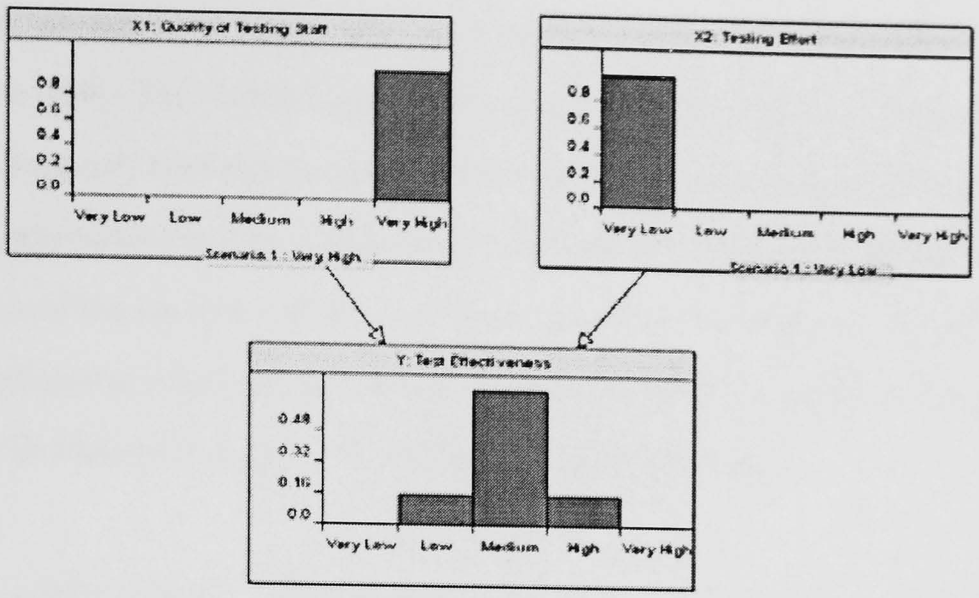


Figure 5.7: WMIN function for Y. "TCAS advise" = "very high", with $W1 = 3$ and "Pilot's Skill" = "very low", with $W2 = 1$

$$WMAX = MAX_{i=1,2,..n}\left[\frac{W_iX_i + \sum_{i\neq j}^{n} X_j}{W_i + (n-1)}\right] \tag{5.5}$$

where the weight $W_i \geq 0$ and n are the number of parent nodes.

In each case, experts need only to supply the parameters to generate the NPT. We found that this set of functions has been sufficient to generate almost all of the ranked node NPTs elicited in practice. The efficiency savings are considerable: if there are m ranked cause nodes, each with n states, then the expert needs to only supply $m+1$ parameter values as compared to requiring $(m+1)^n$ values for full elicitation.

It should be noted that ranked nodes can be further partitioned by declaring additional labelled, Boolean, or numeric parents that can be used to condition the type of weighted expressions that one might wish on the child node.

5.6 Ranked Indicators

We have included this section because, as we explain in Chapter 6, we use this type of relationship to represent the contribution of the organisational culture in company's risk.

We interpreted culture as something that is not directly observable; we can only perceive its manifestations. That is, looking at the performance of a given company, i.e. whether or not it is "successful", we can determine whether it has a "good" or "bad" culture [190]. This culture is reflected on factors like "Unsafe Working Practices", "Workforce Support" that can be observed and thus measured using, for instance, survey responses, performance in jobs and employment turnover. These manifestations represent the *indicators* of the present culture, see Figure 5.8. In this Figure we can see ranked indicators modelling the relationship between "Culture" and factors such as "Unsafe Working Practices", "Workforce Support", "Participative Communication".

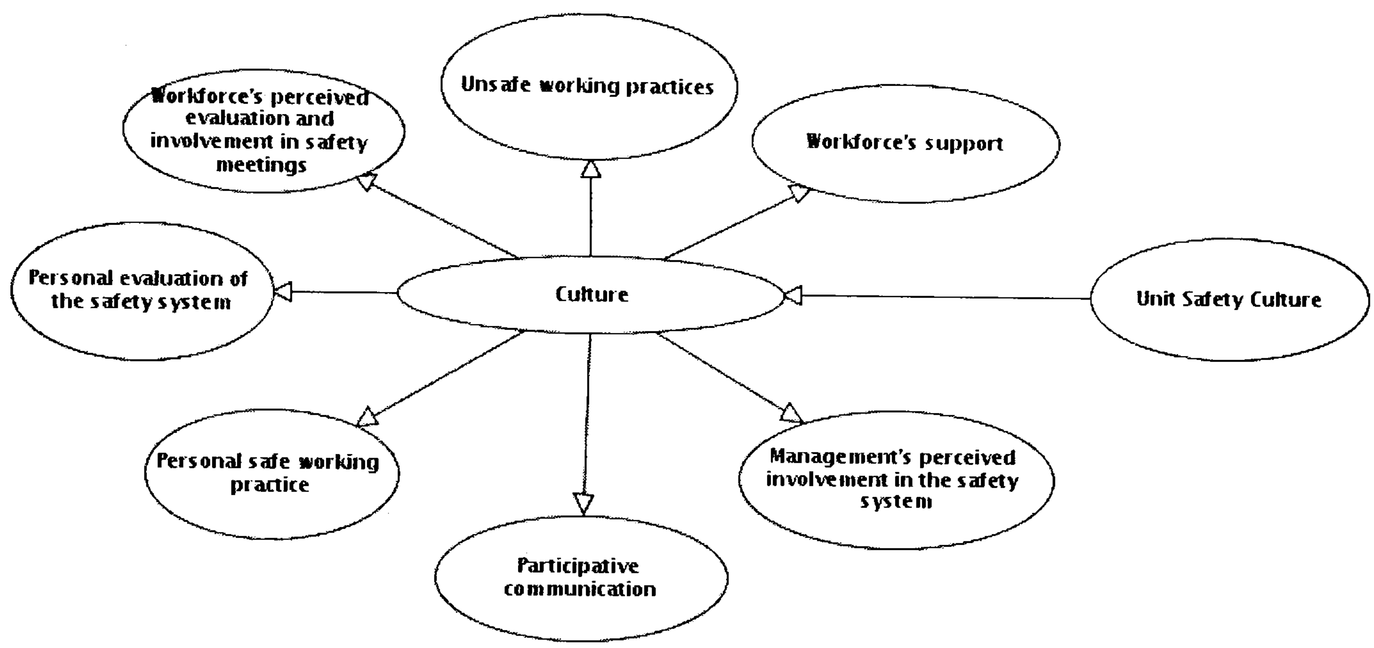


Figure 5.8: Culture sub-net

Indicator nodes operate in a similar way to "filter" nodes in a Kalman filter. Here, we can think of the indicators as providing noisy or imperfect observations and the parent node as the true (but possibly unobservable or not economically measurable) value awaiting estimation [142]. In a Kalman filter, we wish to condition our estimate for the "true" value on the data on hand from each of our "indicator" nodes, assuming that each indicator is Normally distributed.

Formally, the joint distribution for a set X containing $i = 1,2,..,n$ ranked indicators X_i of a single causal parent node Y is

$$p(\underline{X},Y) = p(Y)\prod_{i=1}^{n} p(X_i|Y) \tag{5.6}$$

We model the NPT for each indicator node by using the doubly truncated TNormal distribution:

$p(X|Y) = TNormal(Y, \sigma^2, 0, 1)$

This assumes that the nodes Y and X are on the same scale. The expert simply has to specify the variance parameter σ_i^2 whose inverse acts as a "credibility index": the higher the credibility index, the greater the correlation between the indicator and the parent cause node.

Indicator nodes are correlated with each other following concept of *d-separation* explained in Chapter 3. This correlation is desirable, given that indicators reflect the true state of the underlying unknown cause. Only when the cause itself is instantiated with hard evidence are the indicators uncorrelated. However, given that the causal nodes are usually unobservable (this is, after all, why we use an indicator), indicator nodes are generally not independent in practice.

Another perspective on the use of indicator nodes is that each can be treated either as a different sub-attribute of the parent node or as a different measure of that sub-attribute from a different source, for instance, when there are multiple experts, each with a different credibility, producing different observations.

Part III

Examples

Chapter 6

Modelling Safety of an Air-Traffic Control System

6.1 Preface

This project is a joint effort between the Risk Assessment and Decision Analysis Research (RADAR) group in the Department of Computer Science, Queen Mary, University of London and the Safety Research Unit (SRU), in the Department of Psychology, Liverpool University[1].

In a sequence of five interviews of three to four hours each, over a period of one year, we discussed the model structure, identified and defined the relevant nodes and their relationships. Present from NATS during these sessions were: the safety director, a manager from the Short-term Collision Advise team, two human factors engineers and one air traffic controller (ATC). While the director had some knowledge about BNs, the others had background knowledge in Risk analysis but not in BNs.

During these informal interviews they explained the flow of processes they follow in their daily jobs, identifying the tasks they perform and the tools they use. The group leader, M. Neil, decided earlier in the project to follow an unstructured approach to the elicitation of the model. The idea was to reach a compromise where different views can be accommodated thus providing a framework to take action to improve air-traffic safety, if needed.

We explained to all parties the concept of BN, conditional probability, and the

[1]The SRU has built cultural metrics against which companies can compare their organisational cultural settings.

potential pitfalls during the elicitation process and, most importantly, that there is not a singular model, as there is not a single solution, but a model that best fits their knowledge at a given time. That is, the outcome of a BN model provides an approximation to current knowledge; approximation that given the characteristics of BN (see Chapter 4) will become closer adjusted to the true nature of the event as feedback and data is accumulated.

The elicitation of the probabilities took three more sessions over the same period of time; in two of them the NATS experts drew probability distributions using the ranked nodes approach, and in the last session we compared the model's results against experts' expectations. We had to fine-tuned some of the outputs because they did not match the expected values. This was done for each components of the model. Results were later compared against Safety Significant Events (SSEs) data and presented to the safety director.

The modelling of organisational culture was based on the data collected by our partners in this project the SRU, led by Prof. I. Donnald and his team: S. Johnson and Q. Xie. We organised four sessions of four hours each, two in Liverpool and two in London, where we discussed the network structure. During these sessions we agreed on the definition of culture and its representation in the model.

In the following sections we explain, within the Soft System Methodology (SSM) framework, the development of this project. This methodology as P. Checkland and J. Poulter [31] comment:

> ...was developed using an alternative model of research, ... , namely "action research". The idea of action research is that the researcher will state the framework on which the research is taken place with the purpose of being reproduced, (and thus validated) obtaining similar results (not identical given the human component of the research). The SSM being one of these frames of work.

This methodology is needed because as Checkland [30] argues:

> ...social reality is no refined entity 'out there', waiting to be investigated. Rather it is to be seen as continuously socially constructed and re-constructed by individuals and groups ...

For that reason P. Checkland and J. Scholes [32] comment that

> ...the aim of SSM is to take seriously the subjectivity which is the crucial characteristic of human affairs and to treat this subjectivity, if not exactly scientifically, at least in a way characterised by intellectual rigor.

This is why this methodology is typically used in problem domains that contain a large social component. In this project, this social component translates into the study of the contribution of organisational culture in the avoidance of mid-air collisions.

This contribution, as Checkland [31] would argue, is not a problem as such (i.e. as something that needs fixing) but a problem situation, in the sense that it involves more than one issues as we will see in this Chapter.

It can also be seen as an opportunity for improvement. These improvements are beneficial if we consider the opening of new runways and a consequent increase in air-traffic volume. In this context, we can think of the BN model as a model to question the problem situation and to validate the need to take action, if need be.

Following this methodology, this Chapter is structured as follows: in section 6.2 we discuss the problem situation and in section 6.3 we set the framework of this project. Within this frame we have analysed the activities that are relevant to this problem. In section 6.4 these activities are structured and defined using a CATWOE analysis. One of these activities, Culture, is of particular interest to this thesis. For this reason section 6.5 is dedicated to explain its role on this model. All these definitions, also called root definitions using SSM terminology, provide the bases to build the conceptual model; this is done in section 6.6. Section 6.7 explains how the probability tables were built using the ranked node approach.

6.2 The Problem Situation

Often, the focus of incident investigations in air traffic management is on the human failings of pilots and ATCs. Examples include confusion when reading instruments and failure to recognise or take note of warnings [23]. Human performance, as we discussed in Chapter 2, is influenced by a wide variety of "performance shaping factors", e.g. training, unexpected behaviour of technical systems [205]. These *latent conditions* or "upstream" factors provide the underlying structure, i.e. the chain of conditions, where breakdowns or *active mishaps* occur [190].

The study and measurement of this chain of events allows us to take action in order to improve the safety record of an organisation.

The aim of the NATS Human Resources (HR) team was to investigate this chain of events. Note that we have analysed the air-traffic control processes within the perspective, or in SSM terminology *Wordlview*, of the HR team. The focus of this project is to observe the effect of human factors in the management chain. From this perspective, we have analysed what the HR team and the RADAR team considered, in the words of Checkland [31], *feasible* and *desirable.*

> (Changes)[2] are systematically desirable if these 'relevant systems' are in fact perceived to be truly relevant ... (they are considered feasible)[3] ... only if they are perceived as *meaningful* within that culture ...

For this reason, this model focuses on the activities that are meaningful and relevant to HR. Thus, when we say that a Planner or a Tactical ATC have de-conflicted a potential incident, e.g. loss of separation between two aircrafts, we are referring to the event cease to be a risk threat. The details of how they deconflict this potential risk are not relevant to this model. The same applies if an ATC considers that two airplanes are flying too close and recommends evasive maneuver. The type of maneuver performed by the pilot is not relevant to this model, again, we are only concerned with whether this event was resolved or is still a potential threat. This model does not capture that level of granularity. We are mainly interested in events that contain a human component that can be singled out from the rest of the network, for instance, all the tasks that have to do with routing an aircraft through an airspace corridor can be explained, in human factor terms, as the capabilities of the ATC.

To the question, how close does this model's predictions match their expected outcome, we can say that this model was built following the SSE data. The aim was to match the frequency of events that go through each barrier[4]. The BN model is approximately

[2]Text within brackets added

[3]Text within brackets added

[4]**Note**: All participants of this project signed a non-discloser agreement by which we are not allowed to faithfully reproduce data or previous air-traffic risk models, e.g the British Airways model.

20% - 30% close to their initial estimations. The concept of "barrier" defense will become clearer as we progress through the Chapter.

The next section identifies the relevant activities in the traffic management chain.

6.3 Description of the Problem Situation

The focus of this project is to put into practice, in a real-world model, the ranked-node approach and, in doing so, to observe the effect of the human factors on the management chain and ultimately, by improving this to reduce the risk of mid-air aircraft collision.

This air-traffic control system BN model was developed from a series of workshops with representatives from NATS, the UK's Air Traffic Management authority.

The outline of this model is the effort of Martin Neil (Group leader), Roger Shaw and Bob Malcom (facilitators) and Jose Galan (Technical leader).

Jose Galan, apart from contributing to the development of this model as part of the RADAR team, was personally responsible for building and validating the probability tables and for modeling the cultural elements. The probability tables were built and validated using a computer program Jose Galan implemented for this occasion, see Appendix A. This project was a test bed for the development of the ranked node approach, explained in the previous Chapter. With the help of this computer program, we were able to find, through trial and error, a general approach, i.e. ranked node, that made possible the construction of this model.

The interpretation and design of the impact of Culture in the organisation's safety proved to be challenging. First, it was the novelty of this approach, i.e. a) we are able to measure culture, b) we can define what makes a bad or good safety culture and c) this definition enables managers to take action to improve company's safety record. Second, the design had to be intuitive enough for the participants to understand and reason the outcomes. Third, it had to accommodate the data provided by our partners on this research; the Safety Research Unit (SRU) at Liverpool University.

Note that the development of the Culture network (i.e. human factors having an effect on a firm's safety) is going to be the main thread through out this Chapter. The reason being to show Jose Galan's contribution on this project and, in doing so, to explain his other contributions.

The workshops with the experts at NATS highlighted the following areas of discussion:

- In air traffic management, there is a chain of activities, from design of airspace ("upstream"), through flight planning, through the ATCs and their interactions with pilots, to the reactions of pilots ("downstream").
- These are the activities performed by the ATCs [135]:
  - The design of the airspace.
  - Procedures for controlling the access to, and the use of, the airspace.
  - The use of highly skilled and trained air traffic controllers.
  - The use of collision warning systems, such as the Short Term Conflict Alert system[5] (STCA) and the Traffic Alert and Collision Avoidance System[6](TCAS).
  - The skills of the pilots themselves.
- Each of these activities are intended to minimise the difficulty of operations further downstream in such a way so as to minimise the risk of collision.
- All these activities (i.e. designing, planning and routing aircrafts) are surrounded by the unpredictable weather, aircraft performance and pilot action, any of which might affect the speed or direction of aircraft and therefore upset the plans of the ATCs. Equally, though, such factors might help recovery from a threatening situation.
- There are different types of airspace; from sectors over remote countryside that experience very light traffic (e.g. Scotland) to very busy sectors around major cities (e.g. London).
- There are different types of sectors; some concerned entirely with flights through the sector and some only with takeoffs and landings.

[5]STCA is an automated air traffic control safety net, which alerts controllers of potential conflicts.
[6]TCAS is the equipment aboard an aircraft that gives audible and visual warnings when there is a threat of collision with another aircraft.

- Different teams have different cultures. Different teams, also, typically have different technical systems and may even have evolved different (possibly informal) operating procedures.

- The organisation of the teams is structured differently in different parts of the airspace. In some areas there is the separation between Planner ATC (PATC) and Tactical (TATC) (The role of these two different jobs are explained in the next section).

Within this framework, both RADAR and NATS decided that, in the first instance, a "sector group" should be studied (in this case Swansea), in which teams, usually working in shifts, control traffic through a small number of related sectors. Such sector groups might comprise 5 pairs of PATC and TATC, each pair being responsible (usually) for one sector.

The input of the weather was interpreted as factor that influence all other activities, or in other words, as a contextual probability evenly distributed among the other activities (see section 3.6 in 35). For that reason, it was removed from the discussion on the model.

The output of the model should be a set of expected frequencies of various kinds of breaches of the layers of defense also called "barriers", and that the model should be able to show the extent to which cultural (and other) factors affect those frequencies. This approach is convenient since SSE data is gathered monthly on potential problems which arise through breaches of the various barriers.

6.4 Root Definitions

The starting point to build our model was a previous conceptual model made by British Airways (Figure 6.2 depicts a similar model) and the available SSE data. SSE data is linked to the loss of separation. This data is defined relative to the bands:

- Band 1: Separation <= 66% of prescribed value or, if no prescribed value, then <= 2 nautical miles and 600 feet of vertical separation.

- Band 2: Separation > 66% of prescribed value or, if no prescribed value, > 2 nautical miles and 600 feet of vertical separation.

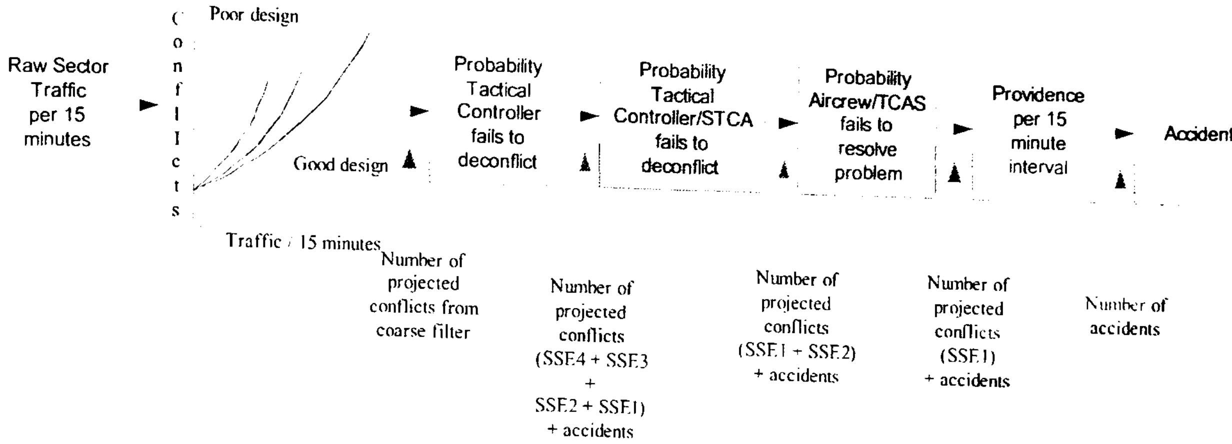


Figure 6.1: Loss of separation data derived from the barriers breaches.

The underlying assumption is that at any given time it is expected that a number of aircrafts will breach one of these bands, Figure 6.1 illustrates this point. The SSE data classifies the event according to the barrier at which the breach occurs[7].

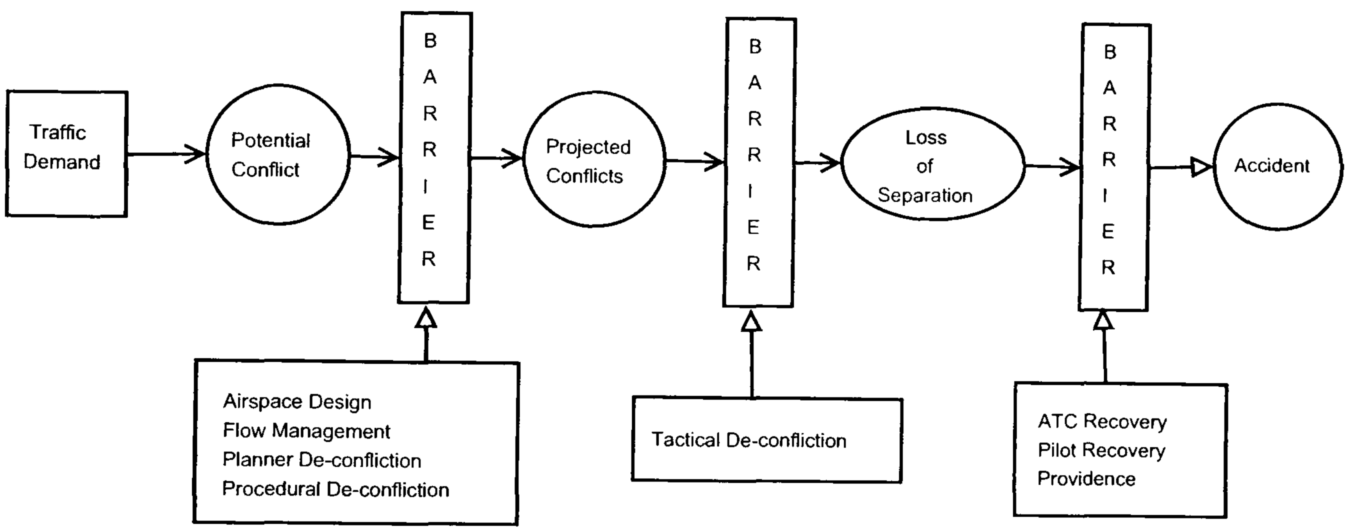


Figure 6.2: The "Barrier Model" of Air Traffic Control.

[7]Classification of SSE data:

- SSEP (prior to Tactical intervention): SSE Potentials occurred for any Band 2 incident.
- SSE4: An SSE4 is registered if a Band 1 event occurred and was effectively resolved by a controller who was providing the service when the event was initiated, and no system or procedure failure affected the resolution. Two sub categories are defined, the first if the controller was aware in advance that the event would occur and the second if he became belatedly aware.
- SSE3: In this case the Band 1 event is detected and resolved by ATC but either it was not solved by the controller providing the service, or the event was detected by another controller, STCA, or the pilot, or it was not resolved in a timely or effective manner, or a system or procedure failure affected the resolution.
- SSE2: In this case, once again Band 1, the pilot resolved the event, or it was not resolved by the aircraft safety net or it was resolved by the aircraft safety net.
- SSE1: Finally, the Band 1 event was not resolved by timely pilot action or there was a high risk that any action taken was ineffective. Matters are resolved by providence.

In Figure 6.2 we observe the workflow of activities performed by ATCs at different barriers, also termed as "defence in depth". All aircrafts are managed, initially, by flow managers. These aircrafts are then subject to sector planning controllers that route them through the airspace. If needed, the tactical air traffic management will handle any potential safety risk. Ultimately, the risk of collision depends on pilot's skills and on what the air-traffic community calls "providence" (or luck).

To help us define these activities we are going to perform a CATWOE[8] analysis. This analysis can be performed at different levels of abstractions:

CATWOE analysis at Project level:

- Customer. Society in general and CAA and NATS in particular. Ultimately, the HR department at NATS in that they, directly, benefit on the understanding of the role organisational culture in the NATS's safety record and on measuring its impact. At the model level, the Customer is the ATC who benefits from the task(s) performed previously down the chain.

- Actor. CAA and NATS acting as the guarantors of the air-traffic safety. At the model level, the actors are the ATCs (or Pilot), considering that they daily tasks consists on the air-traffic management to ensure safety.

- Transformation. Given a volume of air-traffic, with potential risks, as input; the ATCs are responsible to design, planned, route and, if need be, de-conflict potential safety risks thus producing a safe airspace as output.

- Owners. In a wider context, NATS and ultimately the CAA. In a more specific context, we can consider the ATCs and Pilots as the owners, inasmuch as they are responsible to avoid a potential incident going further down the chain.

[8]This analysis consists on defining these activities by studying its constituent or emergent properties.

- Customer, who benefits from the activity;
- Actor who performs the activity;
- Transformation, the conversion the activity undergoes;
- Worldview or the perspective the activity is perceived under;
- Owner who can start or stop the activity and the
- Environment that constraints the activity performed.

- **Wordlviews.** This Worldview is that of the participants during the elicitation of the model. We cannot talk about different Worldviews between ATCs and the HR team given that during the this process there were not major differences in views that could not be discussed and agreed upon.

- **Environment.** This can be subject to different interpretations: the fact that NATS was privatised can be seen as a constraint, equally the fact that participants in this project lacked statistical knowledge. For this model, the environment is regarded as the air-traffic regulation that each activity must fulfill, e.g. aircrafts should be as far as possible, and as economically as possible kept apart.

The following CATWOE analysis concern the different BN modules that make the model. We are going to center our attention on who is the Customer, the Actor and what gets Transformed. We understand that the definitions of Owners, Wordlviews and Environment can be shared. We need these definitions, called *root definitions* in SSM terminology, as the basic building blocks to produce the BN model.

The starting point of the chain of activities is the *design of the airspace*. Conflicts are made more or less likely as a result of the airspace design. This design is taken as given in this model, as the "presenting situation". After this, we have the Flow Management.

CATWOE analysis for Flow Management:

- **Customer.** PATC.

- **Actor.** Flow Management. Typically there will be one or two flow managers for each sector group. They have the task of assigning flight-paths to aircraft. They may or may not be aware of the risks arising from the combination of routes they assign, but air traffic control organisations have little influence over the way flight planning is done, other than through attempts to constrain traffic through flow management.

- **Transformation.** Given a volume of air-traffic, the aim of flow management is to maintain this volume through the various air traffic control sectors within manageable levels (known as "target sector flows"). Aircrafts should be as far as possible, and as economically as possible kept apart.

CATWOE analysis for PATC:

- Customer. TATC.

- Actor. The PATC. Their activities are the routing of an aircraft through a sector.

- Transformation. It separates the airspace into sectors and locates flight corridors through sectors. Maintain aircrafts apart and minimise the risks when aircraft diverge from their planned route - such as at route crossing points between two aircrafts.

CATWOE analysis for TATC. First barrier:

- Customer. Pilots and/or TATC in recovery mode.

- Actor. The TATC has the job of maintaining the separation of aircraft, often through direct interaction with pilots. Also the air-crew may take a greater or lesser part in choosing appropriate avoiding action.

- Transformation. The management of the aircraft through a given route. The TATC would manage the moment-by-moment sector movements and would deconflict potential conflicts using planned procedures. For instance, if two aircrafts may be heading toward a situation where they are too close[9], though not necessarily dangerously close, then there is said to be a "potential conflict". Either the TATC or the air-crew may "deconflict" such a situation by re-routing the aircrafts. Potential conflicts are recorded on the SSE data.

CATWOE analysis for TATC in Recovery mode. Second barrier:

- Customer. Pilots.

- Actor. The TACT people might perform different functions at different times. For instance, an important distinction is drawn between the TATCs working in "no panic" mode - anticipating and avoiding potential conflicts well in advance - and TATCs operating in "recovery" mode, where there has been a "loss of separation" that must be rectified quickly. If there has been a loss of separation then the TATC

[9]Close to 2 nautical miles of horizontal distance between aircrafts or at least 600 ft of vertical separation

intervenes to recover the situation through direct interaction with pilots. They may be assisted by the STCA. Also, of course, the air-crew may take a greater or lesser part in choosing an appropriate avoiding action.

- Transformation. The TATC, generally following standard operating procedures, advises the pilot on an appropriate avoiding action. The TATC may be helped in this task by a warning from the STCA which automatically detects impending loss of separation. If this loss of separation is not resolved, it is recorded in the SSE data.

<u>CATWOE analysis for Pilots. Third and last barrier:</u>

- Customer. All. This is the last barrier of defence. The immediate beneficiaries are the passengers inasmuch as their lives depend on it.
- Actor. Pilots. They may be assisted by the TCAS.
- Transformation. If both aircrafts are in a conflicting situation and have TCAS installed and operating correctly, then these systems should automatically negotiate compatible recommendations for avoiding actions which they present to the pilots. If TCAS is not operating, or if despite TCAS recommendations they find themselves in a collision course, pilots might still with skill and judgement avoid a collision. But if they fail to do so, an accident can happen. This is recorded in the SSE data.

<u>CATWOE analysis for SSE:</u>

- Customer. SSE data.
- Actor. ATCs and Pilots.
- Transformation. Depending on the performance of each of the ATC functions and Pilot's skills, a proportion of a volume of traffic might be expected to "breach" some of the barriers. The amount of breaches at different barriers are recorded as SSE data.

<u>CATWOE analysis for Culture:</u>

- Actor. HR department inasmuch as they are responsible to provide an internal culture in which ATCs work.
- Customer. The ATCs.
- Transformation. The culture of the organisation is transformed. The assumption of this project is that an improved internal culture would translate in a reduction of air-traffic accidents/incidents caused by human factors.

Now, the question is how to examine the culture of an organisation in order to determine whether action needs to be taken to improve it. The next section addresses this point.

6.5 Culture

Modelling "culture" in air traffic management deserves special attention, partly because a fundamental motivation to signal the author's contribution to this model, and partly because it proved to be particularly problematic.

The starting point is the work of the SRU led by Prof. Ian Donald. The SRU research has shown the importance of employees' perception of management involvement in the risk management process. The SRU has produced a Safety Attitude Questionnaire (SAQ) whose answers correlate (to varying degrees) with the actual safety records of a number of organisations. The SAQ dataset was collected from 73 companies across the world during the period of 1992-1999, and includes companies from the oil, chemical and power generation industries. The majority of the respondents were from the UK and also from China (Mainland and Hong Kong), Greece, the Netherlands, Portugal, Australia, Middle East, and New Zealand. The total sample size is 7211 respondents[10].

The SAQ data was broken down into probability tables suitable for inclusion in the BN. To achieve this, it was necessary to rate the individual companies in relation to their safety performance and to provide safety culture data for each of these companies. In an effort to reduce data not related to safety culture in the existing dataset, a number of companies were excluded. The criterion for exclusion was companies with a sample size of less than forty, and companies not based in the UK. Additionally, companies which had

[10]Information provided by SRU team members Q. Xie and S. Johnson

repeated safety culture data (i.e. safety culture surveys completed more than once) were restricted to inclusion on the basis of their first survey only. The final sample therefore included 23 UK companies, with over 4500 respondents[11].

Factor number	12 Factors correlating to the company's safety culture	Correlation
1	Personal evaluation of the safety system	0.90
2	Safety representative's perceived (knowledge of and) involvement in the safety system	0.88
3	Personal safe working practice	0.87
4	Workforce's perceived evaluation and involvement in safety meetings	0.86
5	Management's perceived involvement in the safety system	0.86
6	Unsafe working practices	0.79
7	Safety representative's perceived evaluation of the safety system	0.86
8	Workforce's (perceived safety encouragement and) support	0.80
9	Co-worker's perceived involvement and evaluation of the safety system	0.78
10	Management's perceived evaluation of the safety system	0.70
11	Participative communication	0.70
12	Personal involvement in the safety system	0.70

Table 6.1: Factor Structure of the SAQ. These factors are strong indicators of how "good" or "bad" the safety culture of an organisation is.

There were over 60 questions in the SAQ, that have been grouped into 12 "factors", see Table 6.5. See Appendix B on page 177 for the complete set of questions. From these 60 questions, approximately 20 have been identified as correlating significantly with the company's culture, and for that reason they can be used as strong indicators of how "good" or "bad" a company's safety culture is. For instance, in Table 6.5 we can see the questions related to the first factor: Personal evaluation of the safety system.

Consistent to J. Reason's [190] argument, see Chapter 2, we interpreted culture as something that is not directly observable. We can only perceive its manifestations. Looking at the performance of a given company, i.e. whether or not it is "successful" (e.g. lower rates of accidents, lower turnover), we can determine whether it has a "good" or "bad" culture. In this sense, the causal link between culture and its manifestations forms a loop where the actions reinforce the idea of culture and culture is responsible for those actions. These physical manifestations can be observed and thus measured, e.g. survey

[11]Information provided by SRU team members Q. Xie and S. Johnson

Questions to determine the factor 1: Personal evaluation of safety system
I feel satisfied with the safety information I get
I am happy with the existing safety precautions for particularly hazardous work
I feel satisfied with the attention given to safety in any training I have had
I am happy with the safety equipment specified for my job
Generally I am happy with the safety in my asset area
I know the results of safety inspections to do with my job
The people I work with are satisfied with the attention given to safety in any training they have had
The people I work with are satisfied with the information they get about safe working
If changes are made to the procedures for my job I know about them
I feel I could tell my boss if I had worries about safety

Table 6.2: Questions Correlated to employees' perception of management involvement in the risk management process.

responses, performance in jobs.

Figure 6.3 is the result of this interpretation. The node labeled "Culture" represents "cultural assessment", a probabilistic assessment of a given culture (the "cause"), and the nodes, called *indicator nodes* (see Chapter 5 section 5.6 on page 88), represent the "factors" (the grouped responses to survey questions) that correlate best with safety incident records. The information provided by these indicators classify the culture that an organisation exhibits. Note that this interpretation of culture reflects the ideas discussed in Chapter 2.

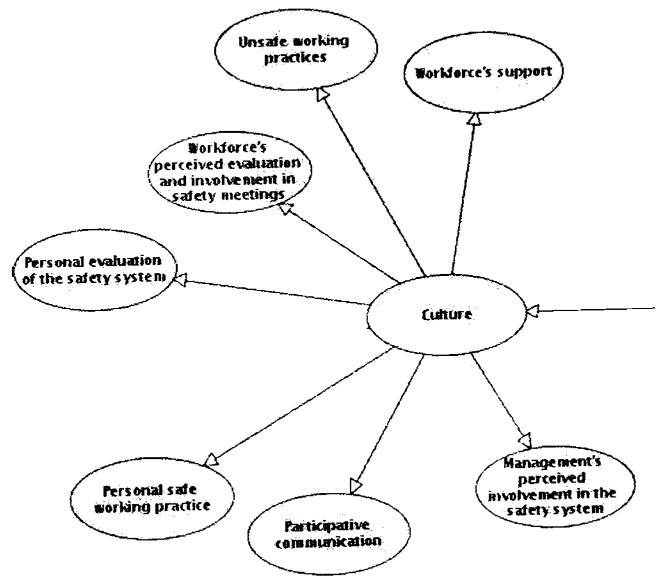


Figure 6.3: Final sub-net for cultural assessment

Also, note that the factors are now reduced to 7. We were concern that 12 indicators nodes would worsen the results rather than to improve them. With 12 factors was more difficult to obtain the results expected by our colleagues at SRU. The SRU team suggested initially that a solution with fewer factors was possible. so further analysis were run forcing data into 7 to 11 factors. We opted to reduced the number of indicators to 7 which was the best compromise between the SAQ data and the indicator nodes solution.

Table 6.3 is a typical example of the data obtained from SAQ. Table 6.4 illustrates how these data was translated into probability tables.

Culture bin	Factor Value	7 Factors affecting organisation's culture						
		1	3	4	5	6	8	11
Very Weak	1	0	1	1	1	1	0	1
	2	0	4	0	1	3	3	3
	3	0	10	4	1	19	2	4
	4	10	11	11	8	15	8	25
	5	23	14	17	23	1	23	6
	6	6	1	5	5	1	4	0
	7	1	0	2	0	0	1	0
Total		40	41	40	39	40	41	39
Missing		1		1	2	1		2
System total		41		41	41	41		41

Table 6.3: SAQ Data for Culture classify as "Very Weak". The Factor Value column range from Strongly Disagree (1) to Strongly Agree (7).

Culture	Factor 1						
	1	2	3	4	5	6	7
Poor	1^{-8}	1^{-8}	1^{-8}	10	23	6	1
...	...	...	...	...	...	...	...

Table 6.4: Conditional probability for factor one given a poor organisational culture. Note that the zero values where changed to 1^{-8} when introduced in the NPTs given that a zero in probability means that a particular scenario is inconceivable. These values are later normalised.

Figure 6.4 is an illustration of the Culture network showing a graphic view of the probability tables.

SRU also suggested to consider just two "levels" of culture that of the "unit" (typically a sector group) and that of the overall NATS organisation.

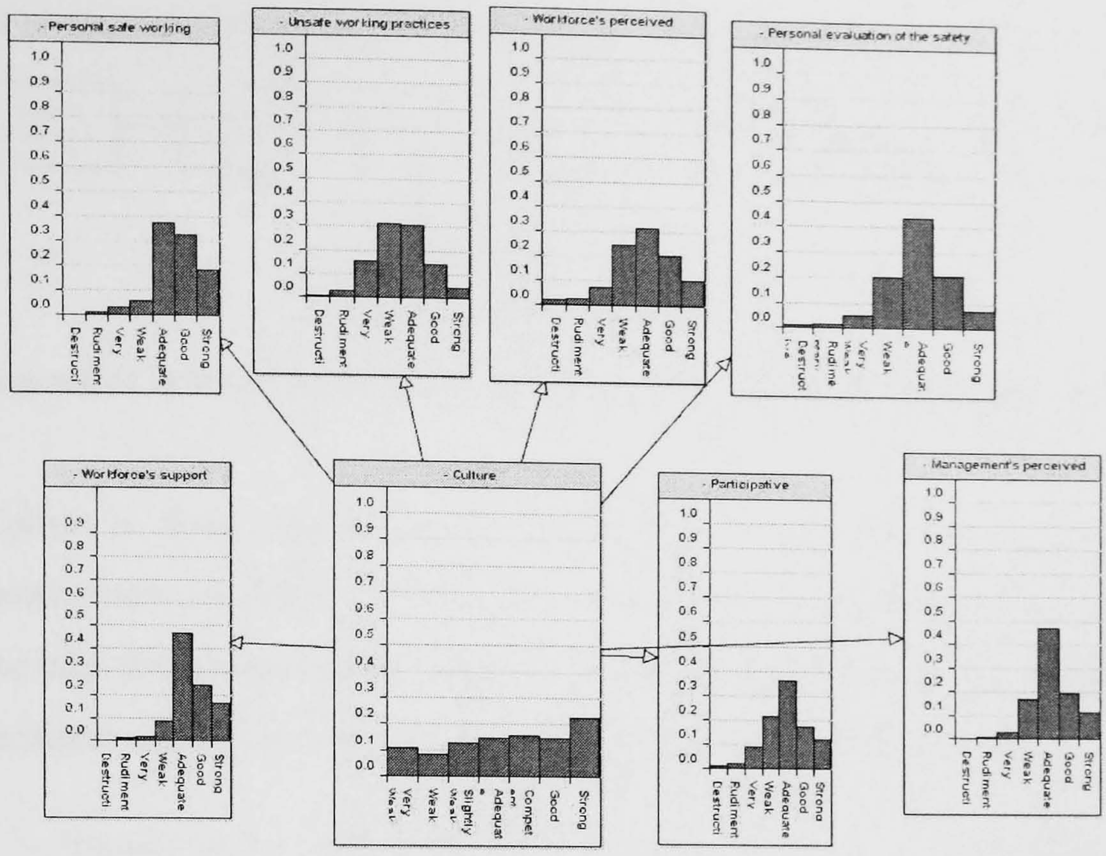


Figure 6.4: Graphic view of the Culture network NPTs using the Agena Risk tool.

6.6 Conceptual Model

At this stage we put together all the previous root definitions to build the BN model. This model is made of different modules:

- The "Culture" or Socio-Technical module. Each socio-technical function might have its own local culture, and each collection of socio-technical functions (as in a sector group) would have its own culture and there would be local cultures associated with each level of the organisation.

  For this reason, this pattern will be replicated whenever an activity has a human component which can be considered as independent, in this case the PATC and the TATC (including when acting in recovery mode).

  The factors affecting the TATCs' capability may well be different for different *modes*. An ATC's performance in recovery mode, for example, is likely to be conditioned by specific training for such circumstances. It is also affected by the performance of particular technical aids, such as the presence and operation of any STCA system. So, in the network the patterns for socio-technical functions were repeated for both "normal" and "recovery" modes of operation of the TATC, treating them as separate functions. Figure 6.5 shows a schema of this interpretation.

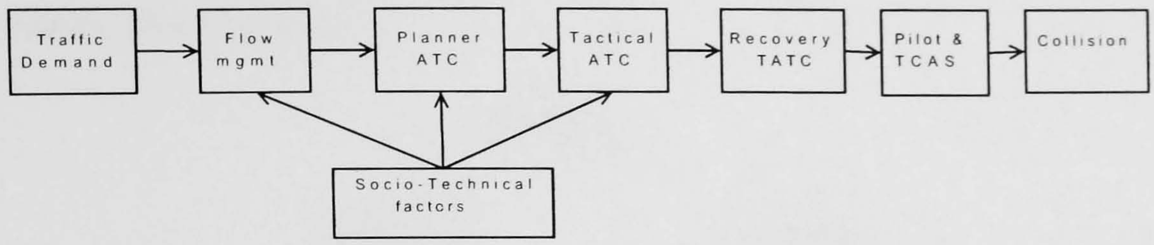


Figure 6.5: Schema representing the role of NATS's culture in the BN model.

Culture is, then, regarded as a factor not just in parallel with "people skills" and "workstation capability", but affecting those factors as well, through such things as the importance attached to training, and the design, maintenance and operation of technical support systems. In Figure 6.6 illustrates this interpretation.

- "People Skills" defines the staff education, training, and experience. Note that this definition does not represent simply competence, which is generally regarded as independent of a given situation, but in this network represents the diligence of application of that competence as conditioned by the local culture.
- "Workstation Capability" refers to the performance and the usability of the systems, as well as the correctness and dependability of their functional design.

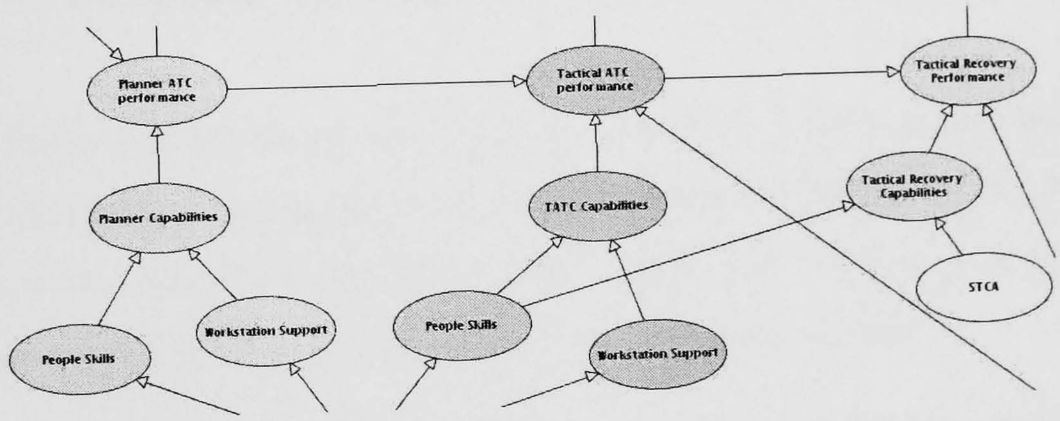


Figure 6.6: Note that the Culture pattern is replicated whenever an activity has a human component which can be considered as independent, in this case the PATC and the TATC (and in recovery mode).

An additional node: "Cultural Dependency", was introduced in order to allow experimentation with the degree of cultural autonomy of the operational units. Often pockets of "unsafe" culture exist within safety conscious organisations; and occasionally "safe" pockets of behaviour exist within organisations with a poor safety culture. See Figure 6.7.

Figure 6.7: Cultural Dependency adjust the impact of the prior NATS safety culture

- Pilot's module. It also was decided to make explicit the combined effect of the pilot (including the aircraft performance and other air-crew members) which affects the performance not of the TATC capabilities, but of the TATC function in which the ATC works with the pilot.

  In addition, the role of the Pilot responding to TCAS warnings is now more clear. TCAS does not itself control the aircraft: its warnings are mediated by the pilot and the aircraft.

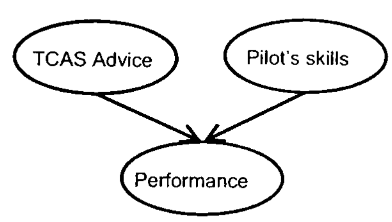


Figure 6.8: Pilot's performance. They are the product of their skill's and TCAS advice.

In Figure 6.8 we can see how the pilot's behaviour is the common cause to the TATC's performance and recovery capability together with the TCAS advice. This means that the pilot's skills have a bigger weight in the outcome of an accident, so much so, that it only needs the pilot to fail to have an accident.

- SSE data module. The role of the SSE data is to fine-tune the outcomes the model will become clearer during validation in section 6.8. Depending on the performance of each of the ATC functions, a proportion of that volume of traffic might be expected to "breach" some of the barriers. We can model this module using a *Markov-chain transition model* [111]. That is, the outcome of a process depends on the previous one, thus making a chain a successive events in time [132]. See top row of nodes in Figure 6.9.

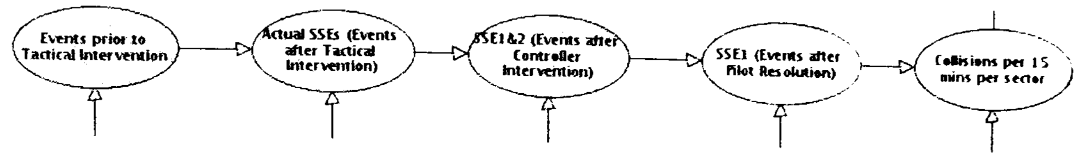


Figure 6.9: SSEs network representing expected barrier breaches.

For instance, a proportion of the volume of aircraft movements handled by the TATC's in normal operational mode will not be satisfactorily "deconflicted" and become potential losses of separation which must be handled by the TATC's in "recovery mode". Ultimately there is only luck (helped by the volume of airspace out there) and then just bad luck and collision.

The network in Figure 6.10 shows the final version of the BN model.

Note that Flow Management is not present in the model. While ultimately it may be desirable to consider how changes to flow management processes might improve "down-stream" performance, NATS decided for the present to focus on the performance of PATC and TATC processes affected by the organisation culture.

So, the outputs of the Flow Management processes have been replaced the input nodes "airspace characteristics" and "traffic complexity" whose output is the "presenting problem" for the "PATC Performance" node. This node aggregates the effects of volume of traffic, complexity of traffic mix, complexity of airspace, such as crossing paths, complexity of flight plans, and the performance of flow managers in trying to keep the traffic presented to PATC's to acceptable levels. See Figure 6.11.

6.7 Populating Node Probability Tables

The prior probability for nodes such as "Traffic Volume", "Traffic Mix", "STCA" or "TCAS" were obtained from the NATS experts.

A more difficult task was to built the NPTs for the conditioned nodes. For each conditioned node, experts were asked to express their beliefs about particular scenarios associated with that node by drawing probability distributions. These curves are estimates of the conditional probability density function for the "output" of a node given its "inputs". Given a selection of scenarios, we are able to generate a function for the child node. After having elicited a full set of graphs, the node probability tables were

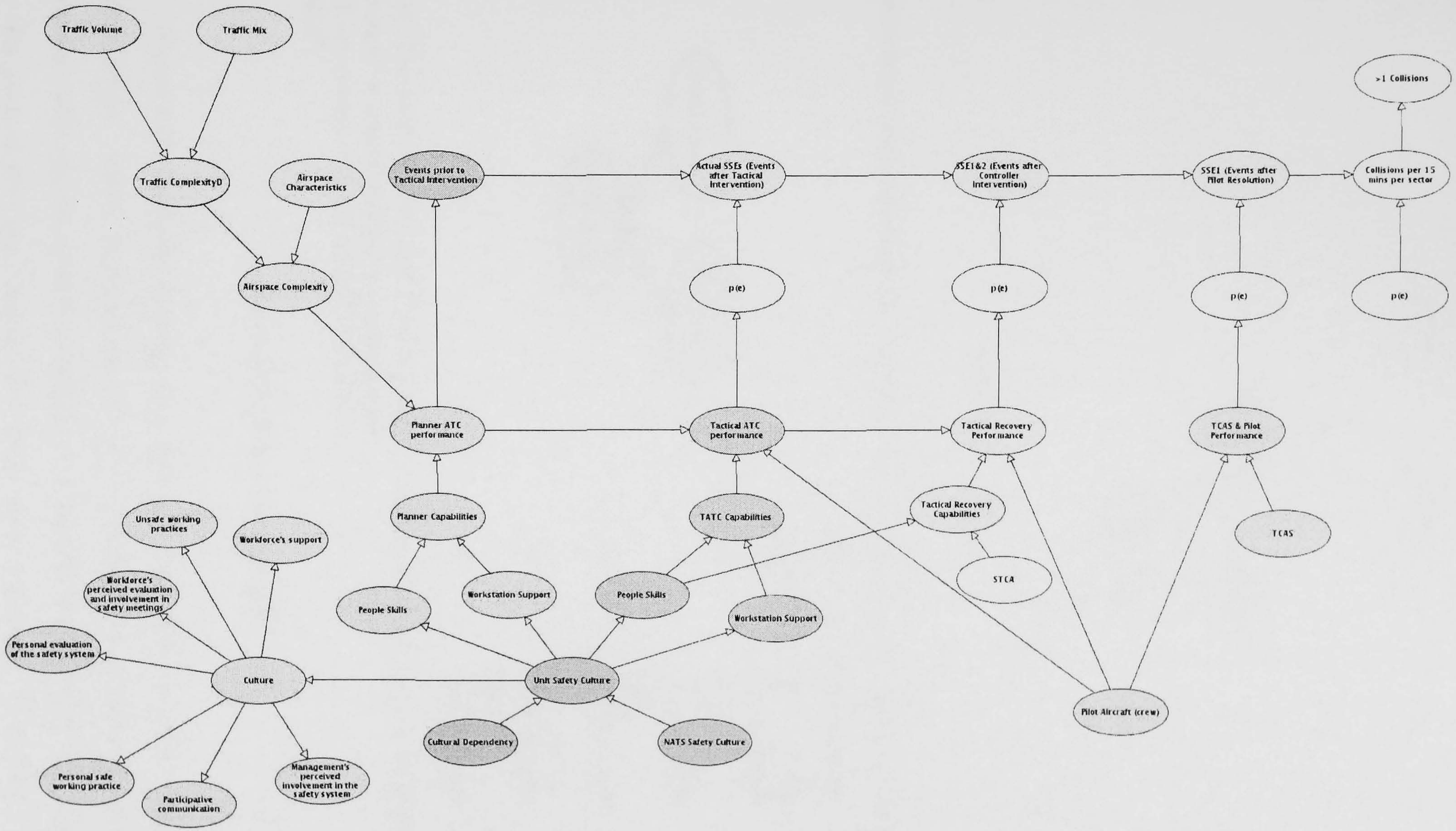


Figure 6.10: Final network arising from the case study. The final network is made of a number of modules.

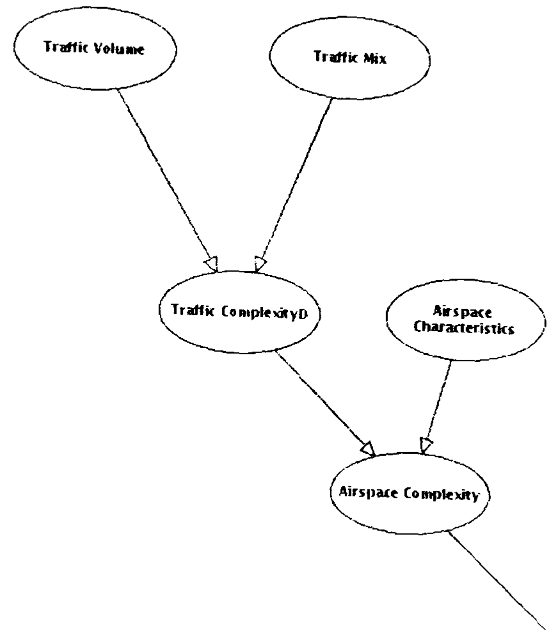
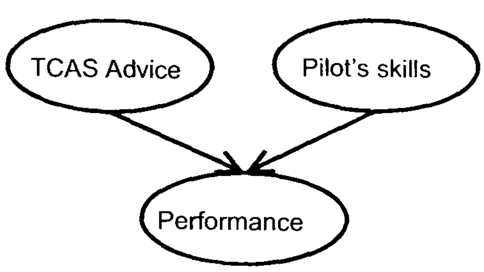
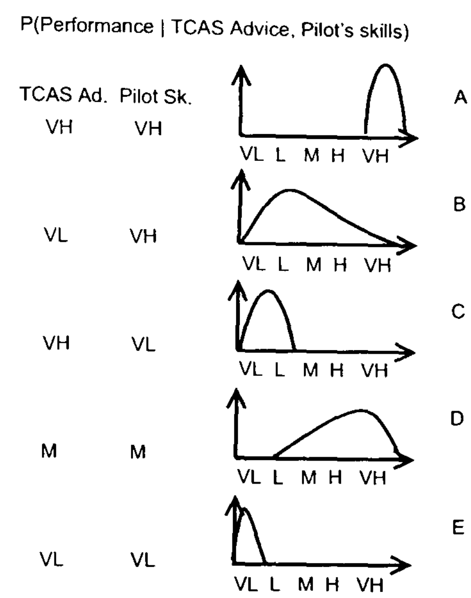


Figure 6.11: Planner

constructed semi-automatically using the ranked node approach described in Chapter 5.



(a) The node "TCAS and Pilot's performance" is considered to be a consequence of two causes: "TCAS advise" and "Pilot's Skill".



(b) Expert's drawings of a set of scenarios.

Figure 6.12: Elicitation example

Figure 6.12 shows the example of a node that represents the pilots' responses to TCAS advise. Those responses depend on two factors - the extent to which TCAS provides useful support, and the performance of the pilot in responding to a TCAS alert.

For each of the input factors, five states were considered: "Very Low", "Low", " Medium", "High" and "Very High".

Instead of eliciting all 125 combinations of these factors, needed for the Performance node, we considered just four extreme situations along with a fifth, medium, case. The experts' probability graphs for these five scenarios are shown in Figure 6.12(b).

The experts explain these estimates as follows: In the case of TCAS supportiveness and Pilot performance both being "Very High" we would expect the resultant Performance to be also very high. A highly skilled pilot may, for instance, add a lateral turn to whatever advice is given by TCAS. Both the "skewness" and the sharp peak of the distribution indicate a high degree of certainty for this estimate. In the case of TCAS supportiveness being "Very Low" and Pilot performance being "Very High" we have a more spread distribution along "Very Low" and "Low" right up to "High", indicating a less certain outcome. On occasions, for instance, TCAS may give a nuisance alert and mislead a pilot. Pilots are generally trained to follow TCAS advice. However, the more skilled they are the more likely they are to have good situational awareness, with knowledge of other aircraft in the vicinity from radio, etc, and will be more likely to respond correctly to the actual situation, rather than blindly following the erroneous advice.

6.8 Validation

For the purposes of validation SSM recommends to compare the conceptual model with the perceived reality [32]. This methodology describes four ways of doing the comparison:

- informal discussion;
- formal questioning;
- scenario writing based on "operating the models";
- and trying to model the real world in the same structure as the conceptual models.

During the development of this model we have exercised all these types of comparison. We use scenario analysis as one of the main techniques to build the conceptual model. This model was made following the structure of the real world process. To validate the outcome we are using SSE historical data as one of the recommended options in SSM.

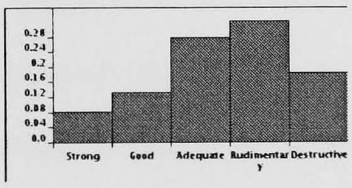
The SSE data is used to corroborate the model's incident estimates, whenever available, and to indicate how often breaches took place. Incident data is currently collected by NATS and is linked to loss of separation, see Figure 6.1.

The SSE data helped us fine-tuned the models' NPTs by comparing the elicited probability estimates with the loss separation data.

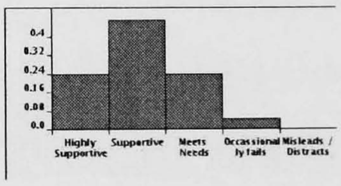
The NPTs for the Culture sub-net, see Figure 6.3, and for its link to the "Unit Safety Culture" node, were agreed with the SRU and were validated later by NATS through the following scenarios.

Note that the purpose of these validation tests is not to compute exact values but to check that the output values appear relatively correct compared with actual experience or the available data. Through scenario analysis, we can observe the relative importance of various factors and their interaction. The following are examples of scenarios we run:

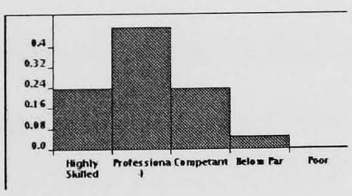
- Default scenario, that is, the model prior to enter any observation. With these settings the network predicts the likelihood of a single accident as 1.4E-07, see right corner of Figure 6.13. This translates into approximately 0.01 events per year on average or as 67.28 years between events on average.



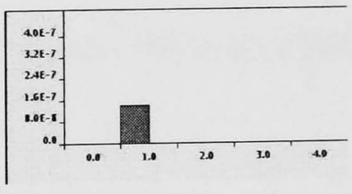
Unit Safety Culture has five states ranging from strong to destructive. The initial node probability table indicates that this variable is as likely to be between adequate (30%) and destructive (18%).



TCAS ranges from Highly supportive to Misleading. With a mean of 46% of being Supportive.



The pilot aircraft node has five states ranging from highly skilled to poor. We observe that around 24% are highly skilled, 47% professional and so on.



Collisions. Probability of an mid-air collision is 0.00000014 on a 15min interval.

Figure 6.13: Default settings

- Scenario 1. This scenario is derived from the default scenario but with the node modelling the sector unit culture set to good. When this evidence is propagated it can be seen that the mean number of collisions per year is 0.01 with a mean of 67.28 years between collisions. The probability of a collision is reduced to 1.65E-08 per year and the probability of the identified SSE events is as shown in table 6.5.

- Scenario 2 models the following node settings:
  - "Traffic Volume" = 30% above target sector flow,
  - "traffic Characteristics" = complex,
  - "ATC Planner" = bellow average and
  - "Unit culture" = good.

  This scenario yields, for example, a mean number of collisions per year of 0.04 with a mean number of years between collisions of 24.98, see table 6.5.

- Scenario 3 models the situation where:
  - "Pilot/crew mix" = below par, (by pilot/crew mix is meant that the capabilities of the pilots, their command of English, their attentiveness to the radio communications, their effectiveness in responding to ATC clearances/requests (how fast do they respond and perform necessary manoeuvres, do they question the need for avoiding action etc.)
  - "TCAS" = highly supportive
  - "Unit culture" = strong

  In this scenario the mean number of collisions per year is 0.03 with a mean number of years between collisions of 35.02, see table 6.5.

Table 6.5 shows a summary of the above scenarios but this time including the SSEs probability estimates.

Scenario	Summary Statistics	All SSEs	SSE1 & 2	SSE1	Collisions
1.	P(event)	1.99E-5	5.75E-06	1.65E-07	1.65E-08
Culture Strong	Mean events per year	17.84	5.16	0.15	0.01
	Mean years between events			6.73	67.28

2. ATC Challenged	P(event)	4.23E-5	1.93E-05	4.45E-07	4.46E-08
	Mean events per year	38.00	17.35	0.40	0.04
	Mean years between events			2.5	24.98
3. Pilots degraded and Mitigations	P(event)	2.16E-5	9.32E-06	3.18E-07	3.18E-08
	Mean events per year	19.40	8.87	0.29	0.03
	Mean years between events			3.50	35.02

Table 6.5: Table with the data predictions obtained from running the above scenarios.

Table 6.6 shows a summary of the scenarios used to validate the model's results, along with the expert's comments on their prediction.

From the validation exercise, it can be seen that the barrier safety process is remarkably resilient to poor performance of particular parts of the process. This is just what one would expect from a "defense in depth".

It might seem that some stages of the process are not very important. For instance, under normal operation, TCAS appears to offer little benefit. However, the important contribution of systems such as STCA and TCAS becomes clear when earlier "barriers" are breached. Again, there is a parallel with redundancy in other reliable systems, redundant parts are only "redundant" when they are not called upon.

Case no.	Scenario	Mean time between events (in yrs)	Commentary
1	We set: Airspace complexity to "Very low" and Planner Capability to "Highly effective"	209	Attention to upstream factors appears to have significant impact on downstream performance
2	Planner ATC Performance: Very low, ATC Performance: Very low and Culture (assessment): good	3	See next scenario
3	Planner ATC Performance: Very low, TCAS: highly supportive Culture (assessment): good and ATC Performance: Very low	11	The poor performance of planners and ATC almost hide the effect of TCAS. In this scenario, TCAS contribution represents a 10%reduction in collisions. (This reduction is not that significant and it was to be expected given that the prior for TCAS is between supportive and highly supportive.)
4	ATC Performance: Very low, Planner ATC Performance: Very low, TCAS: misleading and Culture: good	<1	Now the importance of TCAS can be seen in a poor environment, the situation is ten times worse without effective TCAS.
5	Planner ATC Performance: Very low, ATC Performance: Very low, STCA: misleading and Culture: good	5	STCA has a prior around supportive and highly supportive, so setting this node to misleading represents an increase on collisions to 5 years. Comparing with scenario 4, good STCA appears to offer of halving of the incident rate.
6	Planner ATC Performance: Very low, ATC Performance: Very low, STCA: highly supportive and Culture (assessment): good	10	Consolidates the impression gained in scenario 5
7	Culture (assessment): good, STCA: misleading, TCAS: highly supportive and Pilot: highly skilled	209	Compare with next scenario
8	Culture (assessment): good, STCA: highly supportive, TCAS: highly supportive and Pilot: highly skilled	209	STCA appears to be almost irrelevant when everything else is working well.
9	Culture (assessment): good, TCAS: misleading, Pilot: highly skilled	40	Compare with next scenario
10	Culture (assessment): good, TCAS: highly supportive and Pilot: very poor.	40	These last two scenarios fit well with the beliefs that a skilled pilot can still rescue a poor situation, while a poor pilot is at risk even with good technological support

Table 6.6: Sample validation tests for the complete network

Chapter 7

Modelling Operational Risk

7.1 Introduction

In recent years the financial industry has recognised the importance of operational risk (OpRisk) in shaping the risk profile of financial institutions. This is not only a result of the large losses we mentioned in Chapter 2 but it is also due to new developments such as the use of more highly automated technology, large-scale acquisitions and mergers, deregulation and globalisation of financial services, growth of e-commerce, emergence of banks acting as large-volume service providers and the growing use of outsourcing arrangements [106].

Proof of this concern are the results from the survey conducted in 1997 by the British Bankers Association and Coopers & Lybrand. This survey underlined the importance of OpRisk for the majority of banks; 71% reported an increased concern about it. This is due to the fact that 24% of those banks each experienced losses of more that $1.6 million in the previous three years [11].

For these reasons, the Basel Committee has drafted a system of regulation addressing the issue of OpRisk and its assessment [164]. This Committee is part of the Bank for International Settlements (BIS). The BIS is an international organisation, which fosters international monetary and financial cooperation and serves as a central bank for banks. The focus of BIS is the research and cooperation in international banking regulation.

The key to this regulatory process is the modelling of a business's OpRisk, in terms of a variety of loss event types, in order to arrive at an appropriate regulatory capital charge. The capital charge is the amount of capital banks must put aside as a provision against any possible future losses due to breakdowns. The most important types of OpRisk losses, as the Basel Committee [162] signals,

> .. involve breakdowns in internal controls and corporate governance. Such breakdowns can lead to financial losses through error, fraud, or failure to perform in a timely manner or cause the interests of the bank to be compromised in some other way, for example, by its dealers, lending officers or other staff exceeding their authority or conducting business in an unethical or risky manner. Other aspects of operational risk include major failure of information technology systems or events such as a major fires or others disasters.

In the following sections we are going to discuss the framework set by Basel to control OpRisk's breakdowns. In section 7.2, we give the Basel Committee's definition of OpRisk and study its implications. In section 7.3, we study the regulatory framework. This includes a summary of the regulatory strategies to address OpRisk and an explanation of the Basel proposed methods to calculate the capital charge. In section 7.4, we give a brief introduction of the methods put forward by the financial institutions within the Committee's regulatory framework. Section 7.5 is dedicated to explain the BN approach to model OpRisk. In this section, we decomposed the OpRisk BN model into smaller modules and explain their semantics and probability tables.

7.2 Definition of Operational Risk

In 1998, the Basel Committee [161] came up with different definitions of OpRisk given the different interpretations that financial institutions gave to this risk. For some, it was defined as *the risk not categorised as credit or market risk*, others as *the risk of loss arising from various types of human and technical errors*, among other similar definitions. This initial disagreement in defining the scope of OpRisk highlighted the particular and complex nature of it. The final definition, by the Committee [164] in 2001, is the result of a compromise, wide enough to accommodate these different views. The Basel Committee defines OpRisk as

> The risk of loss resulting from inadequate or failed internal process, people, and system or external events.

This definition signals as causes of OpRisks:

- People. The risk of losses caused intentionally or unintentionally by the employees. It also includes the losses associated with alleged violations of employment law.
- Process. This category reflects losses that have been incurred because of a deficiency in an existing procedure, or of the lack of a procedure being in place. Losses in this category derive from mistakes (i.e. errors) or from not following existing procedures (i.e. compliance or control breakdown). There may be many reasons why the losses occur, like insufficient staff to process the transactions, or because inappropriate access to systems, however, what differentiates the losses in the Process category from those in People, is that the former are not intentional.
- System. This category reflects the risks and losses that are caused by systems/technology. All risks and losses in this category happened by mistake, and are not intentional. If an intentional event occurs, it should be placed in the People category (if by employee) or on the External one (if by a third party).
- External. This category is the risk and losses that occur due to business interruption resulting from natural or man-made forces. Risks and losses that are a direct result of a third party action should be grouped in this category.

Table 7.2 on page 122 shows this classification and the risk event types they can give rise to. It is worth noticing that in a survey conducted by the Risk Management Association (RMA) in 2002 on the sources of OpRisk, the financial institutions consulted pointed out "Process" and "People" categories as the major causes of concern, see Figure 7.1. These results stress the need to implant management practices within the firm to address these potential sources of breakdowns.

7.3 Operational Risk Assessment Framework

The Committee focuses on two areas: (a) the need to establish a strong internal control culture, that takes the responsibility to put into practice regulator's advice, and (b) the

Internal risk			External risk	
People	Processes	Systems	External	Physical
Employee collusion/fraud	Accounting error	Data quality	Legal	Fire
Employee error	Capacity risk	Programming errors	Money laundering	Natural disaster
Employers liability	Contract risk	Security breach	Outsourcing	Physical security
Employment law	Miss-selling/suitability	Strategic risks	Political	Terrorist
Health and safety	Product complexity	System capacity	Regulatory	Theft
Industrial action	Project risk	System capacity	Supplier risk	
Lack of knowledge/skills	Reporting error	System compatibility	Tax	
Loss of key personnel	Settlement/payment error	System delivery		
Employee misdeed	Transaction error	System failure		
	Valuation error	System suitability		

Table 7.2: OpRisk classification. Taken from [164]

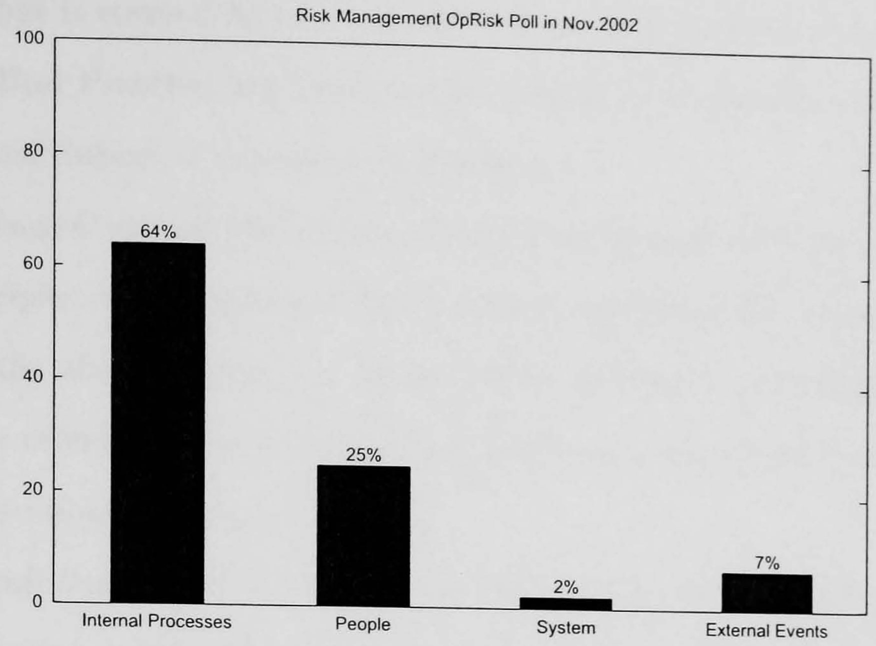


Figure 7.1: Results from a 2002 survey conducted by RMA the sources of OpRisk

need for a quantitative and qualitative method to measure this OpRisk. This amounts to the study of:

1. Corporate governance. Section 7.3.1 gives a brief summary of the regulation and recommendations given to improve the control and assessment of OpRisk.

2. Quantitative and qualitative methods. In section 7.3.2 we study the methods developed within the framework set by the Committee

7.3.1 Corporate Governance - Regulatory Strategies for Improvement

The Basel Committee recommendations echo the concerns and advice that reports such as Cadbury [25], Greenbury Code [195], Turnbull [214] and the Combined Code [54] stated in preceding years. The main difference is that the Basel recommendations focus on banks OpRisk (and Credit risk) and regulate its compliance by means of a capital charge.

All these reports highlight the influence of management practices in shaping the company's risk profiles. They all agreed and emphasised on the importance of the company's directors commitment in the risk avoidance strategy.

The Cadbury report [25], published in December 92, focuses on the role of directors, emphasising the need to implement rigorous reporting and control measures. It also requests the full commitment of the board to risk management, rather than treating it as something that is covered by an insurance policy. The Greenbury recommendations [195] (Code of Best Practice) are based on the fundamental principles of accountability, transparency, and linkage of rewards to performance.

The Combined Code [54] (Principles of Good Governance and Code of Best Practice) is a set of principles, which embraced the Cadbury and Greenbury reports. This is just a summary of the above reports. i.e. board duties, director's control (performance-pay ratio, disclosure of information, remuneration, bonuses, pension. termination of contract payment), financial and internal reporting.

The Turnbull report [143] in 1998 comments on issues such as the type of risks that need to be controlled, keeping the control system up to date and the responsibilities of the board. On this last point, the guidelines are emphatic about the board's responsibility to regularly review the control system [214]:

- To implement the policies adopted by the board;
- To assess reports from management regularly throughout the year;
- To make sure that the control system is working properly, that the system responds adequately to any faults and weaknesses that have been reported and that the corrective actions have taken place;
- Reviews should cover all controls, including financial, operational and compliance controls and risk management;

These above reports consider the environment, where the internal control is exerted, crucial to its success. They consistently agree that the board of directors is the keystone to the implementation of a strong culture of internal controls. This is very much reflected in the Basel Committee's [106] comments:

> The board and senior management should promote an organisational culture which establishes through both actions and words the expectations of integrity for all employees in conducting the business of the bank....despite these differences, clear strategies and oversight by the board of directors and senior management, a strong operational risk culture and and internal control culture, effective internal reporting, contingency planning are

crucial elements to an effective operational risk management framework for banks of any size and scope.

The Committee provides considerable detail on how to achieve this internal culture of control. It does this from different perspectives: from the management point of view [106], providing guidelines to improve internal controls [160–162] or establishing a framework for the OpRisk assessment [164, 166].

In summary, all the reports agree that the internal control system must be integrated within the company's normal management processes and not just as an annual exercise of compliance with the regulators [214]. In essence, OpRisk is about good management, which is not only good for OpRisk stand point, but it is also good for the company as a whole.

7.3.2 Quantitative and qualitative methods - Basel proposed methods

In broad terms, we can classify the OpRisk methods into two groups:

- Top down approaches: These methods estimate an overall capital charge for the firm as a whole.
- Bottom up approaches: The capital charge is based on the measurement of the potential risks within the business units of the firm. This assessment provides more detail on the causes of potential losses.

Within the *top-down* approach, Basel proposes two different approaches: *Basic* and *Standard* and establishes the basis for a third one, *Advanced measurement*, where banks are allowed to produce their own models following either a *top down* or a *bottom up* approach. The Committee's approaches measure OpRisk capital charges in a continuum of increasing sophistication and risk sensibility, (see Figure 7.2). In this figure we observe that the capital charge is related to the managerial "quality". In essence, the firms that demonstrate compliance with the Basel's requirements on managerial internal control practices will have to put less capital aside [160].

The details of the three approaches are:

1. Basic indicator approach (BA). This method is only concerned with the overall capital charge for a firm. That is, is not sensitive to the management strategies to

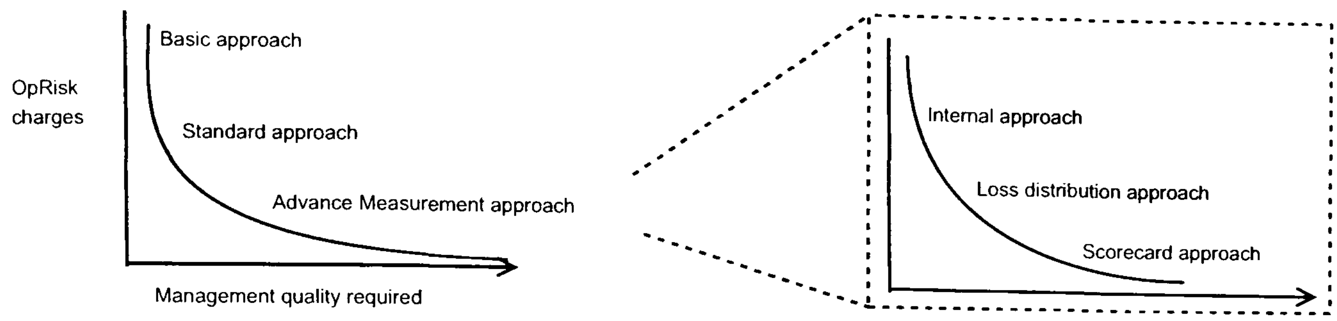


Figure 7.2: OpRisk available methods. The techniques framed with a dotted line correspond to the Advance Measurement Approach.

control OpRisks.

The charge is calculated by multiplying the level of exposure indicator (EI) for the whole institution (provisionally the gross income (GI)) times a fixed percentage α set by the Committee. Following the quantitative impact study in 2003 (QIS3) [167], Basel has set α at 15%. The GI reflects the business volume and thus it can used as an OpRisk EI. According to the provisional definition [33] GI = net interest income + net non-interest income.

The capital charge is, therefore, calculated as follows: Capital charge = $\alpha * GI$.

2. Standardised approach (SA). The SA looks at the potential risks arising from the firm at business unit level. In doing so, the firm can identify and thus prevent the losses. Although SA provides more detail than BA it is still not sensitive to management strategies to control risk. The reason is that SA calculations are not based on internal or external loss data but on the risk coefficients set by Basel.

   In this approach we still use GI as a proxy for OpRisk exposure. The difference with BA is that in the SA the bank's activities are divided into 8 business lines (BLs) and 7 loss event types (ETs) associated with each BL. The Committee [106] has identified these ETs as having the potential OpRisk to result in substantial losses, (table 7.3 in page 128 shows the definition and gives examples of the ETs). The result is a matrix of 8 BLs times 7 ETs as the one shown in table 7.3.

   The total capital charge is the addition of the charges of each BL.

   Capital charge = $\sum_i \beta_i * GI_i$ where i=BL, and GI is taken as EI.

   The percentage, β, charged depends on the BL's risk level. The percentage will vary from a 12% of the GI for the least risky BLs like retail banking to 18% for the

	Internal Fraud	External Fraud	Employment Practices and Workplace Safety	Clients, Products and Business Services	Damage to Physical Assets	Business Disruption and System Failures	Execution, Delivery, and Process Management	Total Across Event Types
Corporate Finance	4 0.0%	3 0.01%	16 0.06%	15 0.05%	8 0.03%	1 0.00%	33 0.12%	80 0.29%
Trading and Sales	16 0.06%	6 0.02%	37 0.14%	112 0.41%	10 0.04%	39 0.14%	1,114 4.07%	1,334 4.87%
Retail Banking	593 2.17%	7,798 28.49%	579 2.12%	1,273 4.65%	837 3.06%	570 2.08%	6,807 24.87%	18,457 67.43%
Commercial Banking	93 0.34%	1,180 4.31%	55 0.20%	66 0.24%	285 1.04%	474 1.73%	1,463 5.35%	3,616 13.21%
Payment and Settlement	22 0.08%	961 3.51%	9 0.03%	57 0.21%	40 0.15%	64 0.23%	752 2.75%	1,905 6.96%
Agency and Custody Services	6 0.02%	7 0.03%	12 0.04%	69 0.25%	17 0.06%	11 0.04%	356 1.30%	478 1.75%
Asset Management	4 0.01%	4 0.01%	21 0.08%	35 0.13%		6 0.02%	360 1.32%	430 1.57%
Retail Brokerage	7 0.03%	2 0.01%	12 0.04%	122 0.45%	28 0.10%	291 1.06%	609 2.22%	1,071 3.91%
Total Across Business Lines	745 2.72%	9,961 36.39%	741 2.71%	1,749 6.39%	1,225 4.48%	1,456 5.32%	11,494 41.99%	27,371 100.00%

Table 7.3: Number of Individual Loss Events per BL and ET. 30 Banks Reporting Data. Taken from [105]

Business Lines (BLs)
Corporate finance $\beta_1 = 18\%$
Trading and sales $\beta_2 = 18\%$
Retail banking $\beta_3 = 12\%$
Commercial banking $\beta_4 = 15\%$
Payment and settlement $\beta_5 = 18\%$
Agency and custody services $\beta_6 = 15\%$
Asset management $\beta_7 = 12\%$
Retail brokage $\beta_8 = 12\%$

Table 7.4: Business Lines β coefficients

most risky ones for example trading. For other lines such as commercial banking a figure of 15% is used [166]. These coefficients are shown in table 7.4.

3. Advance measurement approach (AMA). The methods in this approach follow a *bottom up* approach (apart from the Scorecard approach that can also follow a *top down* approach). The AMA approach allows banks to develop methods to assess their own OpRisks. Financial institutions must be able to demonstrate that the risk measure used for regulatory capital purposes, reflects a holding period of one year and a confidence level of 99.9%, see figure 7.4. Leaving only a 0.01% chance of suffering catastrophic losses. The incentive for banks following the AMA is to

Event Types			
Level 1	**Level 2**	**Definition**	**Examples**
Internal Fraud	Unauthorized activity	Losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law or company policy, excluding diversity/discrimination events, which involves at least one internal party	Transactions not reported Unauthorized transaction Mismarking of position
	Theft & Fraud		Fraud, theft, extortion, embezzlement, robbery, malicious destruction of assets, check kitting, impersonation, insider trading, ...
External Fraud	Theft & Fraud	Losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations by a third party	Fraud, theft, robbery, , check kitting, forgery
	Systems Security		Hacking damage, theft of information
Employment Practices & Workplace Safety	Employee Relations	Losses arising from acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity/discrimination events	Compensation, benefit, termination issues Organized labour activity
	Safe Environment		General liability, employee health and safety rules events
	Diversity & Discrimination		All discrimination types
Clients, Products & Business Practices	Suitability, Disclosure & Fiduciary	Losses arising from an unintentional or negligent failure to meet a professional obligation to specific clients (including fiduciary and suitability requirements) or from the nature or design of a product	Fiduciary breaches, guideline violations, suitability/disclosure issues, retail consumer disclosure violations, breach of privacy misuse of confidential information, lender liability
	Improper Business or Market Practices		Antitrust, improper trade/market practices, market manipulation, insider trading, unlicensed activity, money laundering
	Product Flaws		Product defects, model errors
	Selection, Sponsorship & Exposure		Failure to investigate client per guidelines, exceeding client exposure limits
	Advisory Activity		Dispute over performance of advisory activities
Damage to physical Assets	Disaster and other events	Losses arising from loss or damage to physical assets from natural disaster or other events	Natural disaster losses (earthquakes, fire and floods) Human losses from external sources (terrorism, vandalism)
Business Disruption & System Failures	Systems	Losses arising from disruption of business or system failures	Hardware, Software, Telecommunications, Utility outage/disruptions
Execution, Delivery & Process Management	Transaction Capture, Execution & Maintenance	Losses arising from failed transactions processing or process management, from relations with trade counterparties and vendors	Miscommunication, data entry, maintenance or loading error, missed deadline or responsibility, model/system misoperation, accounting error/entry error attribution, delivery failure, collateral management failure, reference data maintenance
	Monitoring & Reporting		Failed mandatory reporting obligation, inaccurate external report
	Customer Intake & Documentation		Clients permission missing, legal documents missing/incomplete
	Customer/Client Account Management		Unapproved access given to accounts, incorrect clients records, negligent loss or damage of client assets
	Trade Counterparties		Non-client counterparty misperformance, misc. non-client counterparty disputes
	Vendors & Suppliers		Outsourcing, vendor disputes

Figure 7.3: Event types that BLs can give rise to. Taken from [105]

reduce their capital charge percentage to a minimum of 9% [118]. Although this is not the primary aim of the Committee, for them the goal of the institutions following this approach is not to downsize the capital charges but to *right-sizing* it [33].

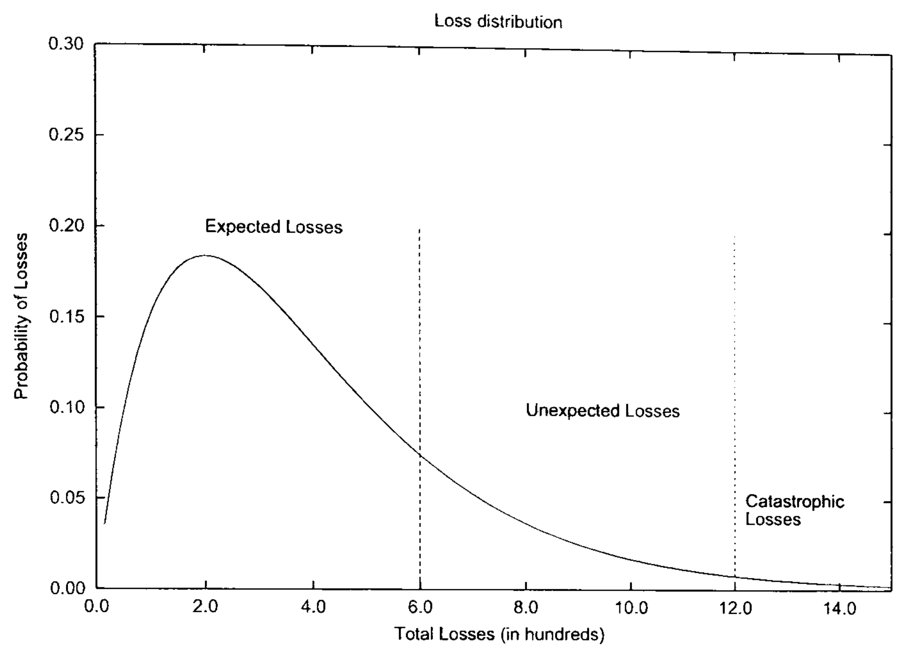


Figure 7.4: OpRiskVaR. The mean of the distributions represent the expected losses while the tail represent the unexpected ones. The losses that go beyond the 99.9% are called catastrophic. Note that similar Figures are shown in a number of books related to this subject, see [8,33,43]. Jose Galan has created this Figure using GNUPLOT graphics tool. This Figure was also used in [153]

The AMA methods must include certain key features such as internal and external data, scenario analysis and factors reflecting the management environment and internal controls [166]. In this context, the Basel Committee [106] has recognised as possible tools for identifying and assessing OpRisk:

- Self-Assessment: using, for instance, a Scorecard approach a bank can assess its operations and activities against potential risk vulnerabilities by ranking the different types of OpRisk exposures. This subjective measurement is not only able to complement internal and/or external data but it also captures the quality of the internal controls.

- Risk Indicators (RI): These are statistical measures that can provide an insight into a bank's risk position. Such indicators may include the number of "failed trades", "staff turnover rates" and "frequency and/or severity of errors and omissions". RIs can be seen as a combination of subjective and objective information. They are subjective because each institution may choose different RIs, given each of them conduct their business in different ways, to gauge their risk levels. They are objective in the sense that these indicators, often, provide statistical metrics, e.g. number of failed trades [8].

- Measurement: this refers to the use of a bank's internal and/or external historical loss data to compute the charge. The resulting charge is calculated aggregating the frequency and the severity distributions.
- Risk Mapping: In this process we identify the factors that are relevant to potential losses and how they interact following causal reasoning. The quantification of these factors can be done either using historical data, expert's judgement or both.

Table 7.6 on page 131 gives a summary of the above techniques

7.4 Advance Measurement approaches

In this section we explain the different techniques developed, thus far, within the AMA framework by the financial institutions. It is worth pointing out that new techniques are coming up constantly which reflects the complexity and novelty of the task. Accordingly, the Basel Committee is making progressive changes to the regulatory draft to accommodate financial institutions' feedback. For this reason rather than explaining each technique in great detail we are going to give an overview of the trends they follow, i.e. whether they are based on data and/or on expert opinion. Authors such as Alexander [8], Cruz [43], Hoffman [100], Chorafas [33] provide a more detail account of the AMA techniques. Figure 7.7 classifies the current techniques depending on the source of information used, i.e. subjective or objective or both.

7.4.1 Data approaches

1. Value-at-Risk. VaR is defined as *the maximum loss over a target horizon such that there is a low, pre-specified probability that the actual loss will be larger* [114]. We are referring to the market and credit risk techniques used to calculate the portfolio volatility. These techniques were, and still are in some cases, the starting point for calculating OpRisk, given the familiarity that financial institutions have in

Approach	Capital charge	Percentage	Source of Information	Comments
Basic	$\alpha * GI$	α=15%	Gross income (GI) GI = net interest income + net non-interest income	The advantage of this method is that is simple and uses readily available data, i.e. GI. Banks do not need to invest money on databases or new techniques to assess OpRisk. The α coefficient is given by the Committee. Thus, there is no incentive for the institutions to improve risk management strategies. In this sense is not risk sensitive. Backward looking, in the sense, that OpRisk calculations do not incorporate subjective information.
Standard	$\sum_i[\beta_i * GI_i]$	β varies from 12% to 18% depending on BLs	GI per BL and ET	This approach inherits the characteristics of the BA. The main difference is that it reflects better the different risks across the BLs. Do not need to collect historical loss data. It uses the GI per BL as a guide for potential losses and a β coefficient proportional to the level of BL's risk. However detailed, it still does not look at factors such as managerial support or staff performance. Thus, regardless the quality of internal controls a bank may have, the charge remains the same. In this sense, it is not sensitive to any risk avoidance strategies.
Advanced	The method used must demonstrate OpRisk value at 99.9% in one year holding period	It can be as low as 9% depending on whether the financial institution can demonstrate compliance to Basel requirements	Quantitative: Using financial's internal loss data or from external financial institutions. Qualitative: Self-Assessment, RIs, Causal Mapping, Scenario Analysis	It aims at promoting the development of new techniques to calculate OpRisk. It takes into account bank's control systems inasmuch as it uses historical data and subjective information. It requires an investment for the building of databases or on new techniques to assess OpRisk. Forward looking inasmuch as it incorporates subjective information.

Table 7.6: Basel proposed approaches

Subjective (Expert's judgement)	Objective (Historical data)
- Self-Assessment: : e.g. Questionnaires, check lists, Scorecards - Scenario Analysis: "What if scenarios"	Measurement: - Value-at-Risk - Loss Data Approach. (e.g. Actuarial models, Extreme Value Theory) - Internal measurement approach (IMA)
- Risk Indicators, e.g. Key Performance and Key Risk indicators - Risk Mapping, e.g. Bayesian networks	

Table 7.7: AMA methods. Classification based on the source of information.

using VaR methods to calculate unexpected values. Although the Basel Committee recognised that this tool is not the appropriate one it is also acknowledged the lack of a better alternative [118]. There are a number of reasons for not using VaR methods:

- OpRisk is essentially different from market and credit risk. As the Basel Committee [106] recognises

  > ..it is clear that operational risk differs from other banking risk in that it is typically not directly taken in return for an expected reward...At the same time, failure to properly manage operational risk can result in misstatement of an institution's risk profile and expose the institution to significant losses.

- VaR methods require data that is available for the market and credit risk but in the case of OpRisk this may not always be true.
- VaR methods calculate the portfolio volatility over a period of time that can vary from 10 days holding period to a month, in contrast to the 1 year period required by the OpRisk.
- VaR does not say anything about the origin of the losses. It just indicates the probability of a value occurring. It treats the losses as inherently volatile. Thus, it does not motivate the management to take a pro-active action to avoid risks.

- The default confidence value for VaR methods is 95% while for OpRisk it is 99.9%. On this point, Jorion [114] comments that VaR methods would not be able to measure risk at such a high level of confidence.
- VaR methods are more concerned with the severity of the losses than with the frequency of those which from the point of view of managing risk is not ideal situation [43].

In the remainder on this thesis we are going to refer to OpRiskVar for the value-at-risk for the OpRisk to differentiate it from the VaR in the market and credit risk.

2. Internal Measurement Approach (IMA). Following this approach, banks must classify their business units in BLs then measure, based on their internal and external loss data (LE), the probability of loss event (PE) and the loss given that event (LGE). The expected loss per BL is given by the product of PE, LE and the EI of that line per ET across the BL [163]. Therefore, the expected losses on a BL is

$$Expected\ loss_{BL_i} = \sum_{j=1}^{j=7} EI_j * PE_j * LGE_j \text{ where } j\text{=ET and } i\text{=BL}$$

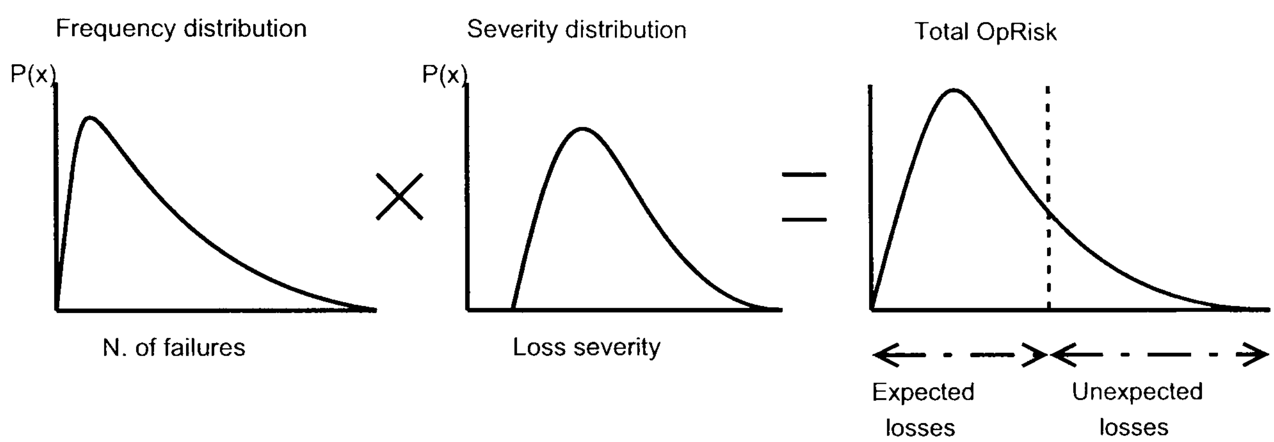
The total capital charge is the summation of the expected loss per BL times the BL's *gamma factor* γ across the BLs (the γ factor is the risk coefficient associated to an ET in a BL). The unexpected losses is obtained multiplying the expected losses by the *gamma factor* [164]. Which means that in this approach unexpected losses are causally related to the expected ones.

$$Capital\ charge = \sum_{i=1}^{i=8} \sum_{j=1}^{j=7} \gamma_{i,j} (EI_{i,j} * PE_{i,j} * LGE_{i,j}) \text{ where } i\text{=BL}$$

It is worth noting that the Committee requires the addition of individual risks (i.e. ETs per BL) unless the bank can demonstrate that the ETs are correlated [184]. The gamma coefficients $\gamma_{i,j}$ are established by the Basel Committee (based on the report data gathered from a number of financial institutions). These parameters are subject to regular validation based on subsequent loss experience or other techniques.

Chorafas [33] points out that the main difference between IMA and SA is that the former provides a more detailed account of the business risks, that is, it can take twenty or more business channels and a dozen or more OpRisk ET, thus providing a more flexible and detailed approach to measure BLs losses. However, this also brings the complexity of calculating the gammas γ by BLs per each ET.

3. Loss Data Approach (LDA). This approach is an application of actuarial methods that combine a frequency distribution describing the occurrence of operational losses and its associated severity distribution. Actuarial models use loss data to build the frequency and the severity distributions of the losses. The frequency distribution describes the number of losses over a fixed period of time and the severity distribution describes the size of the loss once it occurs [43]. Figure 7.5 illustrates this point. The total losses are the result of the frequency of the losses times their severity.



[!h]

Figure 7.5: Loss distribution

This approach deals with unexpected and expected losses separately. It aims to assess unexpected losses directly rather than assuming a causal relationship between them, as the IMA approach does [164]. Thus, the capital charge is based only on the unexpected losses. However, we argue that the reason for this separation is more practical than theoretical. The expected losses are generally provisioned by the financial institutions through insurances. For this reason, the Committee has agreed to lift the charges coming from the expected losses and focus only on the unexpected ones, provided that the institution can demonstrate that it has established a strong internal controls that cater for the expected losses [8]. If the Committee does not separate both risks the financial institutions may end up paying twice for the same risk.

One of the methods used to calculate unexpected losses directly is *Extreme Value theory* (EVT). We are are going to give a brief summary of this theory. Our aim is to put the EVT in the context of OpRisk and not to provide a detailed description of this theory[1]. The use of EVT in estimating financial extreme exposure is relatively new. This theory has been mainly used in the field of engineering (in particular reliability theory) and to forecast extreme meteorological conditions such as floods and tornadoes.

EVT focusses on the extreme values of the distribution, i.e. the tail of the distribution representing large losses, and ignores the rest of the values of the distribution, i.e. the mean. EVT forecasts unexpected losses by extrapolating extreme past events to future events [43]. The unexpected losses refers in this context to *greatest* historical losses experienced by the financial institution. The forecasting is done by fitting parametric distributions to these extreme observations (which, by definition, are a small set of observations) and obtaining the parameter estimates that produce the best fit. There are a number of distributions available to calculate extreme values according the their level of kurtosis, e.g. from Weibull and Gumbel to Pareto distributions to name but a few [33, 147]. There are two approaches to calculate EVT [43].

- Generalised Extreme Value distributions (GEV). GEV is a combination of the following three distributions:
  - Fréchet
  - Gumbel
  - Weibull
- Generalised Pareto distribution (GPD). These distributions are built on the losses that exceed a threshold, e.g. Peaks-over-threshold (POT) distribution. Expected and unexpected losses are separated given a threshold. The setting of this threshold is arbitrary and thus particular to each bank [119].

The GEV distributions are defined by three parameters $F(x;\xi,\mu,\sigma)$. Where μ and σ are the location and scale parameters, respectively. The parameter ξ defines the tail of the distribution [43], for $\xi < 0$ we have a light-tail distribution like Weibull, for $\xi = 0$

[1]For further reading on the subject we recommend Castillo et al. [63] and Christoffersen [35]; both provide examples.

a medium-tail distribution like Gumbel and for $\xi > 0$ a heavy-tail one like Fréchet. That is, the ξ index is proportional to the level of kurtosis. Equation 7.1 shows the cdfs for the GEV. Cruz [43] explains two different methods to estimate the distribution parameters: *probability weighted moments* and *maximum likelihood*.

$$F(x;\xi,\mu,\sigma)\begin{cases} exp\left\{-\left[1-\frac{\xi(x-\mu)^{\frac{1}{\xi}}}{\sigma}\right]\right\} & for\ \xi \neq 0 \\ exp\left\{-exp\left[\frac{(\mu-x)}{\sigma}\right]\right\} & for\ \xi = 0 \end{cases} \tag{7.1}$$

Figure 7.6 shows an example of the EVT using a Gumbel distribution. Here we suppose that a financial institution forecasts a maximum possible loss of around 10.000 for the current year. This prediction is, however, uncertain and therefore subject to variation. The institution wants to find out with 99.9% confidence the probability of suffering major losses. In this case, mode=10.000 and scale between 1 and 2, we give this range to capture the uncertainty in the forecast. The confidence at p=99.9% can be obtained using equation 7.2.

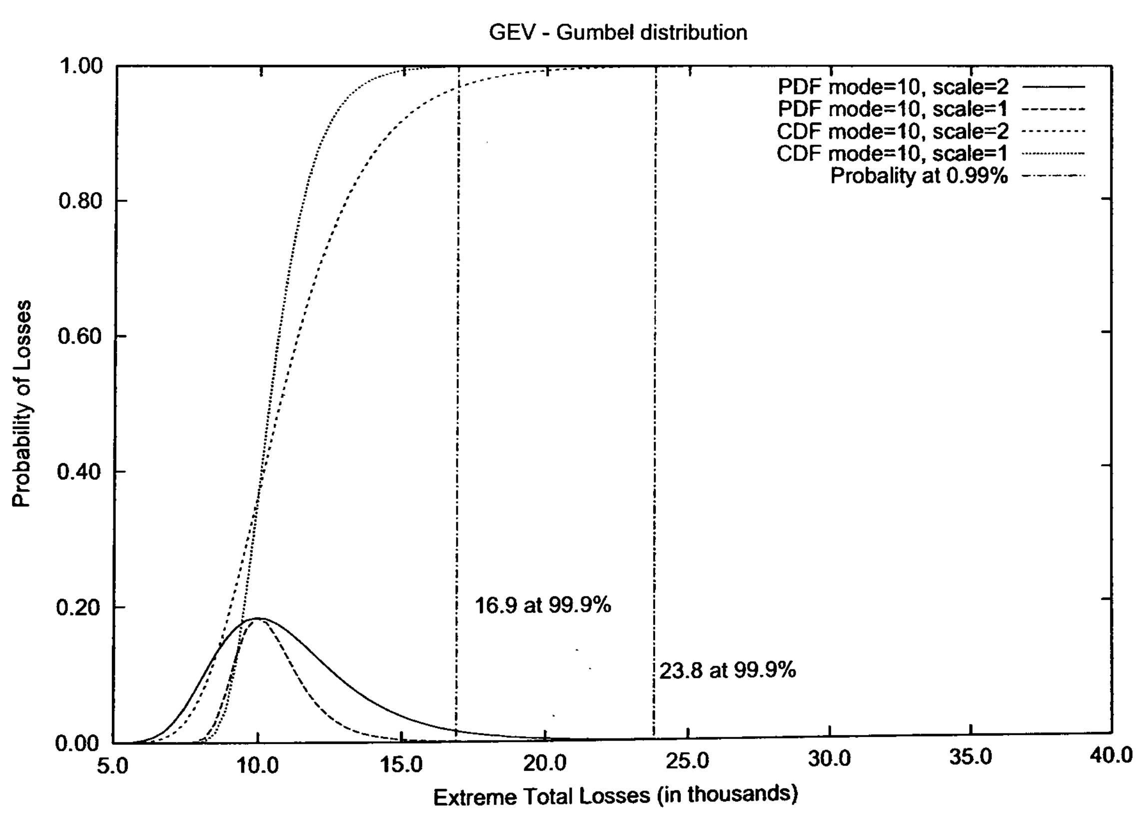


Figure 7.6: GEV - PDF and CDF Gumbel distributions

$$G(p) = mode - scale * log\left(log\left(\frac{1}{p}\right)\right) \tag{7.2}$$

However, the use of EVT in the field of OpRisk has been criticised. For instance, Chorafas [33] questions its applicability given the EVT assumptions, i.e. the loss distribution must be smooth without discontinuities or spikes.

The techniques reviewed, thus far, all rely on data. However, as we discuss in the next section, this brings a number of problems.

7.4.2 Limitations with data approaches

The Basel committee has acknowledged the need to complement historical data with qualitative information. Although the Committee does not explicitly state it, expert opinions must be taken into account to model OpRisk if the Committee wants financial institutions to produce reasonable models by 2007, which is when the OpRisk regulation becomes into action. Implicitly, the proposal of tools such as Scorecards or RIs to assess OpRisk, recognise this need. In our view, there are a number of reasons why purely historical data approaches are not viable:

- Sparse data. It is only recently that financial institutions have started gathering OpRisk loss data. However, this data must extend beyond 5 years to be a statistically significant. Given the time, the data may represent varying levels of operational effectiveness and risk/threat level. Thus, it would be difficult to capture potential losses from one single distribution with a small number of "known" parameters [153].

- Changing environment. Losses experienced are simply a sample of possible events. They may not be representative of changing operational processes. As the underling operational process degrades or improves the value of such historical data, loses its predictive value.

- Under-reporting. We must also look at the context where the data, when available, is extracted. In the domain of OpRisk, a crucial part of building databases is the employees' willingness to report the required information. However, as we know from the discussion in chapter 2, employees are reluctant to give up such information for fear of retribution. For example, consider what Kalhoff et al [118] says regarding the building of a common international database of reported financial losses:

> ... the results from the latest Loss Data Collection Exercise (LDCE) conducted by Bank for International Settlements confirms this impression (the lack of reported frauds) ... the losses derived from the category "Internal Fraud" has only a 3.3% of the total loss events in the LDCE.

This seems somehow unexpected if we think that the main aim of building such a database is to reduce "Internal Fraud" which, unreported, is a major concern for the bank's board (as the Baring's Bank bankruptcy showed us).

It is worth noting that the Chicago based insurance company, Aon, calculated the mean size of bank's fraud to be around \$3.5 million while the Basel's quantitative study (QIS3) [167] concluded that 98% of losses through fraud were for sums less than 1\$ million [45]. According to the data compiled by Aon [45],

> ...fraud is a far greater operational risk than banks have been prepared to admit (in public).

- Difficulty in building large databases. The building of large databases poses a problem. On one side, to share the data banks must collect it using the same standards to identify and classify the events. On the other, banks should be free to select whichever method they think appropriate to assess their risks, and this may entail a different approach to identify and classify risks [33].

- External data. The Committee recognise the use of external data from consortiums such as *OpVantage*[2], *British Bankers Association*[3] or the *Operational Risk-data eXchange Association*[4] run by ABM-Amro, BNP Paribas, JP Morgan Chase, Canadian Imperail Bank among others. Their purpose is to create and share an OpRisk database between its members [8,33]. However, we argue that using external loss data gathered from other institutions is subject to the same problems we

[2]See http://www.algorithmics.com/solutions/opvantage/index.shtml
[3]See http://www.bba.org.uk
[4]See http://www.orx.org/

commented on earlier and may be more because very often the provenance of this data is unknown. Different banks may have different procedures to gather data or different managerial approaches to risk management, e.g. banks setting different thresholds on what they consider a reportable "near-miss". That is, this threshold depends on the severity of the incident which in turn depends on the bank's size, thus, raising doubts about the scalability of the data.

- Assumption of independence. The condition for the validity of data collected by different external financial institutions is that the set of observations among institutions must be independent. However, this assumption is questionable, as data collected during 2001 and 2002 by the LDCE, after the September 11 terrorist attack, shows [147].

- Classification of risk. The loss data can produce an unrealistic prediction given that banks may be double-counting the same incident, e.g. Credit risk can be categorised as OpRisk or a risk from one BL can be assigned to another. As the survey conducted by PricewaterhouseCooper [186] in 2003 points out,

  > .. 69% of banks stated concern about the quality of the data regarding the overlapping with other risks, i.e. credit and market risks.

  Thus, Stone [210], vice-president of risk management at Zions Bank corporation in U.S., comments:

  > Many businesses are reluctant to base their capital allocations or risk improvement efforts exclusively on loss data, as confidence in the data is questionable.

  The British Bankers Association echoes these concerns when it believe that [33]

  > .. that past operational risk loss data does not enable prediction of future losses.

For these reasons many banks are reluctant to invest (around 40% of banks consulted [90]) on databases that haven't got a proven record, given that databases are based on a balance between costs and returns.

- From the point of view of management Turnbull [113] comments

> ..(building) cumbersome risk databases can be a distraction from the primary goal of getting each person in the organisation to be aware of, and manage risks related to the tasks that person performs.

As the study on bank's assessments on OpRisk, in 2001, conducted by the consultory firm TCAS in London [90] concludes

> Banks are realising that it's not about populating a database or about implementing a Monte Carlo simulation and getting an answer. It's about process analysis and identifying drivers ...

These limitations with loss data approaches mean that traditional statistical methods are unlikely to provide useful predictions of operational losses. Therefore, in our view, we need an approach that combines qualitative and quantitative methods. That is, OpRisk models will have to rely on a variety of input data: subjective (e.g. self-assessment and scenario analysis), objective (e.g. loss data) or both (e.g. risk indicators, causal networks).

The Fitch survey [159], in 2004, of over 50 major world banks corroborates this view. It signals that the world's largest financial institutions are adopting a combination of processes to identify risks: 65% are using risk and control self-assessments, more than 32.5% use key risk indicators as the main tool to identify OpRisk. Some 37.5% use risk mapping and 10% scorecards as additional tools.

7.4.3 Subjective approaches

7.4.3.1 Self-Assessment - Scorecards

Following this approach, each business unit or the firm as a whole (i.e. *bottom-up* or *top-down* approaches), identifies the nature and size of OpRisk using a "Scorecard" approach [106, 164].

Simons [207] gives us an example of how to carry out a *top down* Scorecard assessment. In an article, in the Harvard business review, titled "How risky is your company?", he applies the Scorecard system in the context of a "risk calculator". The basic idea is that by identifying the "key pressure points" and giving a score to those points, say, from

1 for low risk to 5 for high risk, managers can work out the total risk exposure of the company. He signals a number of risk indicators:

1. Fast-growing business, e.g. aggressive company's strategy to gain market share may lead employees to by-pass company's policies to get greater revenues.

2. Rate of expansion in operations, e.g. company infra-structure not following the pace of growth.

3. A high level of inexperience among staff, e.g. recruitment policies can be ignored when companies need staff quickly to cope with growth.

4. Company's culture, e.g. internal competition. Do I share information that can benefit others?

5. Information flow in the company, e.g. do failures get reported?

6. Information management, e.g. more sophisticated products brings new risks

As Simons, points out, there is no specific score that firms have to reach. Each firm has its own way to carry out business, i.e. controls and risks are assessed in the context of meeting specific business objectives. What the Scorecard does is to make managers aware of the risk involved in the business activity or process by highlighting the strengths and weaknesses of existing controls.

Using Scorecards in the context of a "bottom-up" approach would consists for instance of filling the matrix 7.3 in page 127. In this process, employees and managers who are responsible for a particular business function present their views about processes, controls and risks quantifying those with a score [50, 119, 123].

The fact that Scorecards do not need to gather data to assess company's risks, makes it possible to capture environments that can change rapidly, for instance, given new management, new products or new technology. Following an only data approach would need years to gather enough information. It also is able to capture organisational factors such as management culture, and system support.

With this approach Basel intends to reflect the improvements in risk management and to bring a forward-looking element into the capital charge calculations. Basel requires a strong quantitative basis for the Scorecards. In this respect it is similar to the requirements for the IMA and LDA approaches. The difference is that this approach relies less exclusively on historical loss data in determining capital charge. In essence, the lack of data must be balanced with strong internal controls that can demonstrate the validity of the measures taken to avert losses.

7.4.3.2 Scenario Analysis

Employees are asked to estimate future possible losses by conditioning them to a particular scenario. Performing this analysis helps us understand the interaction among the variables in the model. Based on these estimations, financial institutions can put capital aside to cater for the risks arising from those scenarios.

We gave some examples of this approach in chapter 6 during the validation process of the NATS model, section 6.8. Then, we performed "what if" scenarios and "stress testing" to obtain the probability of an air-traffic incidents under specific conditions. During the stress-testing we condition the variables in the model to situations that could cause extraordinary losses [114].

7.4.4 Combining Subjective and Objective approaches

7.4.4.1 Risk Indicators

Vinella [221] makes a distinction between risk indicators (RIs): (a) Key Performance Indicators (KPI) which assess the risk of achieving a firm specified level of performance and (b) Key Risk Indicators (KRI) which assess the risk involved in achieving such performance. The idea is that in order to achieve the targeted performance the company must identify and assess the potential losses. In the context of OpRisk, we can think of KPI and KRI as the factors that better explain the operational losses [220]. As King [119] comments

> ..the performance measure is used to develop associated risk measures.. the reasoning is that if performance measures reflect the cause-and-effect relation between the business activities and the earnings of the firm, then a measure of volatility for the performance will reflect the cause-and-effect relation between business activities and risk of earnings.

Certain KRI such as audit scores, staff turnover, trade volumes, error rates, income volatility and so forth are used to measure the risk changes over time. These indicators are assumed to correlate with a change in OpRisk. Therefore, they are used to gauge the company's risk level. The mapping of potential allows the bank to analyse the causes as well as to link the consequent financial losses [201]. This way the risk management process can take proactive steps to reduce losses. This mapping is similar to Faulty Tree Analysis, common in the field of engineering and safety systems where we focus on the different ways a process can fail. This mapping consists of

- Find the potential breakdowns,
- identifying in the organisation the breakdown originates and what parts are affected
- and measuring the severity of the breakdown.

The only requirements that the Basel Committee [201] set, regarding KRIs, is that they should be:

1. Relevant to the frequency and severity of the losses.
2. Non-redundant, if two KRIs are strongly correlated, only one should be considered.
3. Measurable. Objectively quantify and verifiable.
4. Easy to monitor. Ratio between benefit and cost.
5. Auditable.

One method used to measure the OpRisk using key indicators is statistical regression [221], another is using Scorecard as a tool to convert qualitative risk assessment into quantifiable risk indicators [100].

7.4.4.2 Causal mapping

Following causal reasoning we identify potential risks associated with the processes and assess their respective severity. Using BN models we are able to map process workflow while identifying possible errors in the procedures or/and in their execution or/and in the associated controls.

In chapter 5 we explained how we could use our *ranked node* approach to elicit probability values in a similar fashion as the Scorecards. We also showed how we could use linear regression coupled with a Truncated Normal to produce NPTs. When data is available, these estimations can be complemented with historical loss data.

A BN model, therefore, captures all the Basel [166] requirements to qualify for an AMA. That is, a BN can use historical loss data, make use of KRIs to highlight process weakness and causal reasoning to understand their interaction and potential threats, use Scorecards to quantify expert's opinion in changing environments where data is absent thus providing a forward-looking component in the capital charge calculations and is able to perform scenario analysis to corroborate its forecasts. The next section explains the OpRisk BN model we developed within the AMA framework.

7.5 OpRisk Model

The aim of this model is to provide a measure of OpRisk that takes account of qualitative and quantitative factors and causal process relationships. The main assumption that our BN model takes is that unexpected losses are explained as a sequence of "expected" breaches. As explained in chapter 2, major unexpected losses are the result of the accumulation of minor incidents. That is, the unexpected losses are proportionally related to the expected ones. Which means that both types of losses can be represented by the same statistical distribution. The mean of the distribution represents the expected losses while the unexpected ones are captured in the tail.

Our model assess the capital charge at BL level, that is, we have divided the bank's business activities into BLs. The total loss is the addition of all the losses per BL. So in total there would be 8 models representing the 8 BLs. We have built the model for the *Trading and Sales* BL, thus providing a template for how to develop the other BLs models. The tenet is that a BL's model is made up of modules common to all BLs and of modules specific to each BL. For instance, the *Trading and Sales* BL has different KRIs than the *Retail Banking* one, e.g. shares dealings and credit card fraud, respectively. But it also has common elements such as socio-technical factors i.e. staff performance or company's culture, however different these common components impact the BL's losses. Hence, for each BL, we only need to model a specific module representing the KRIs of that line and change the NPTs of the common modules given that each BL has a different level of performance and risks.

We also assume that the frequency and severity are correlated. It is reasonable to think that both would ultimately depend on the organisation's internal controls. The frequency of the potential losses are a reflection of the management efficiency and the same can be said about the severity. With this last assumption we obtain the conceptual model in Figure 7.7. The OpRisk model is made up of the following modules:

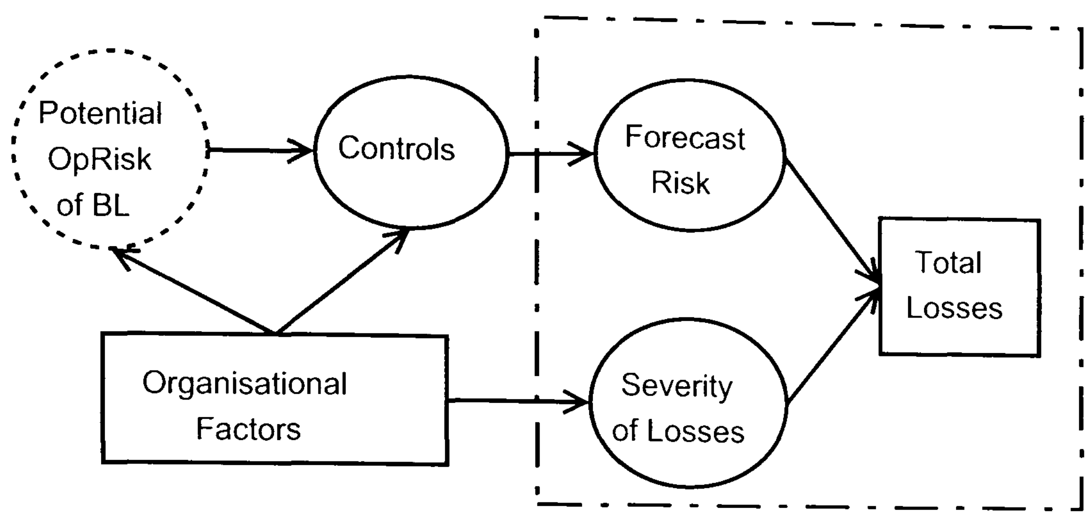


Figure 7.7: Conceptual OpRisk model. Potential losses module is specific to the BL. The total losses is equal to the aggregation of the severity and frequency functions.

- Potential OpRisk of BL. This module represents the losses for a specific BL.
- Controls. Understood as the bank's controls to avoid potential losses.
- Organisational factors. This represents the underlying firm's structure where OpRisks occur and controls measures are exerted.
- Total losses. This is the aggregation of the BL's forecast risks and their associated severity.

The model in Figure 7.7 was further refined to include the BL's EI and the controls associated to the BL (e.g. manager supervision) which are different to the ones exerted at firm's level (e.g. auditing). The rationale of introducing the EI is that the number of events depends on the volume of business of the BL. Figure 7.8 shows the final OpRisk schematic model. In this figure we observe the following modules:

- Exposure indicator. BL's GI.
- Potential OpRisk of BL. Represents the losses for a specific BL and their associated controls.
- Organisational factors. Same as above.

- Controls. Financial control's at firm's level.
- Total losses. Same as above.

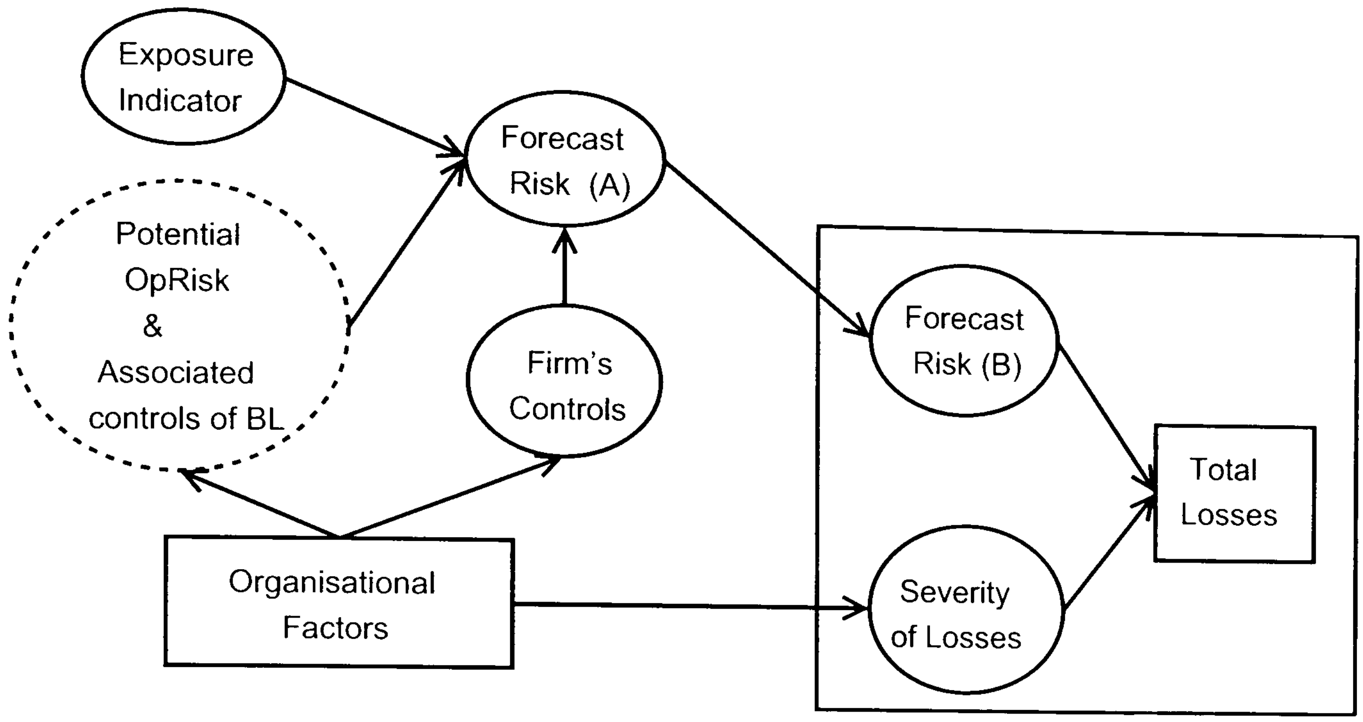


Figure 7.8: Final OpRisk scheme. Forecast risk(A) represents the result of the BL's potential losses given the BL's EI. Forecast risk(B) represents the probability of those potential losses not being captured by firm's controls.

The following subsections explain these modules in detail.

7.5.1 Exposure Indicator

This model takes the GI of the BL as the EI. The GI represents the volume of capital for that BL. This is what Basel [164] has defined as the net interest income + net non-interest income. The potential number of OpRisk events is proportional to the GI of that BL. For example, a BL with a greater number of transactions is more at risk of delaying, mistyping or more likely to suffer fraud.

7.5.2 Potential OpRisk and Controls of the BL. Key Risk Indicators.

Using BN models we can map business risk indicators and show how they interact to create a potential risk. In this context, risk is interpreted in a wider sense: risk is understood as a threat of monetary losses and failing to achieve the potential profits, e.g. a business transaction can become a threat if it is delayed for too long or bring profits if it is dealt with efficiency.

Potential losses emerge when KRIs acts in conjunction with other KRIs to create or increase the likelihood of an event, e.g. if there is no clear segregation of duties and the trading is not well supervised the likelihood of losses due to fraud increases. We have modelled this interaction using "OR" and "AND" gates, see Figure 7.9.

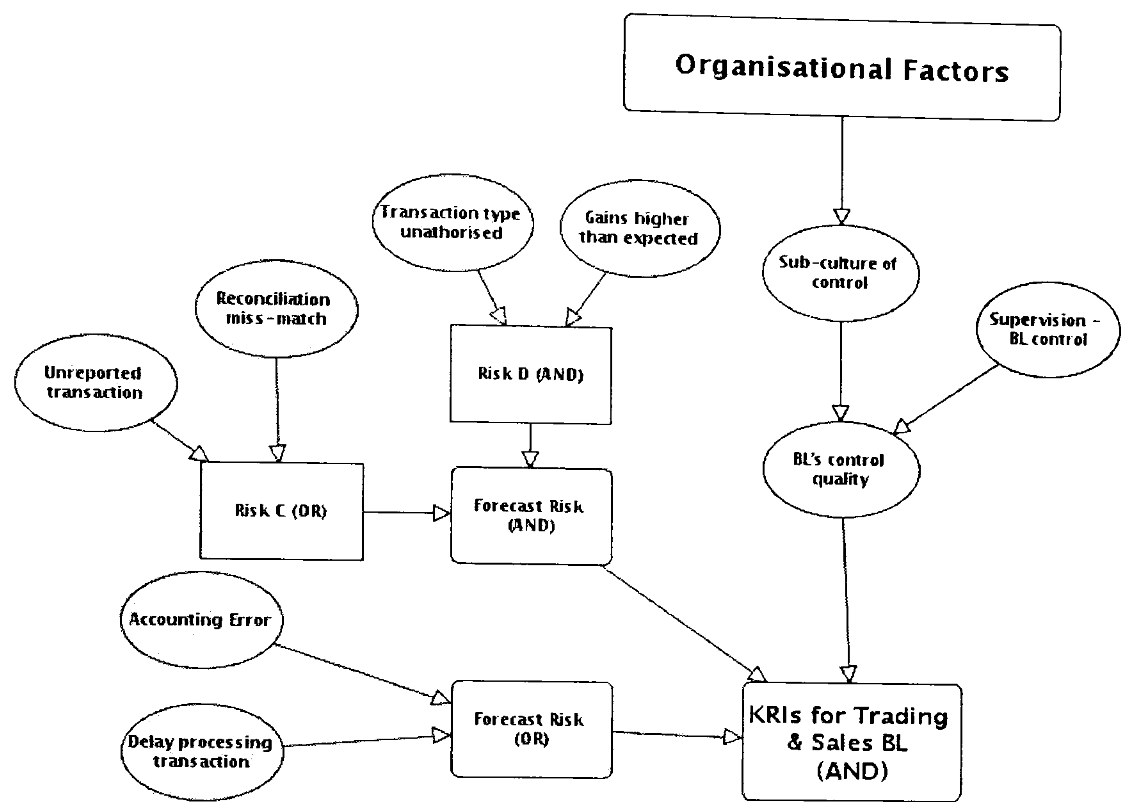


Figure 7.9: KRIs mapping. The nodes showing risks indicators interaction through "OR" and "AND" gates.

In an "OR" gate node we are concerned with the smallest probability of a risk emerging while in an "AND" gate node we are only concerned with the highest probability value. Notice that the gate nodes in this model, labelled with an "AND" or "OR", capture the interaction of both parent nodes but they have not got any other causal meaning attached to them.

Figure 7.10 shows an example of the "AND" gate. In this case, we want to model the interaction where either of the events, "Transaction type unauthorised" and "Gains higher than expected", has the highest probability value given that any of them can cause great losses for the firm. In this figure we see that we have evidence on the node "Transaction type unauthorised" being "Very High", see Figure 7.10(a). This evidence makes the shape of the distribution for the child node "Risk D" to move towards "High". Notice, however, that when we introduced new evidence on the node "Gains higher than expected" being "Very Low" the probability distribution for the node "Risk D" is not altered.

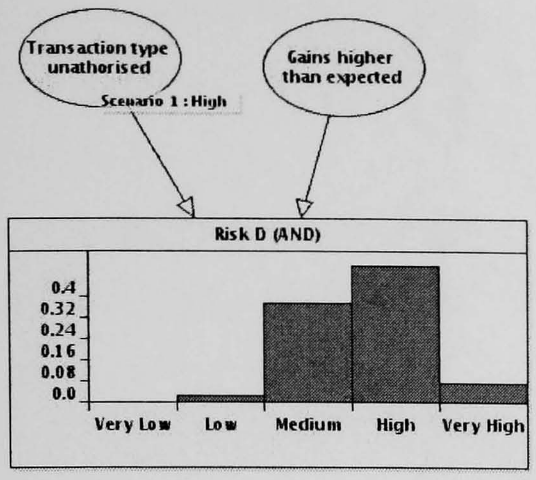


Figure (a). The probability of a "Transaction type unauthorised" is "Very High".

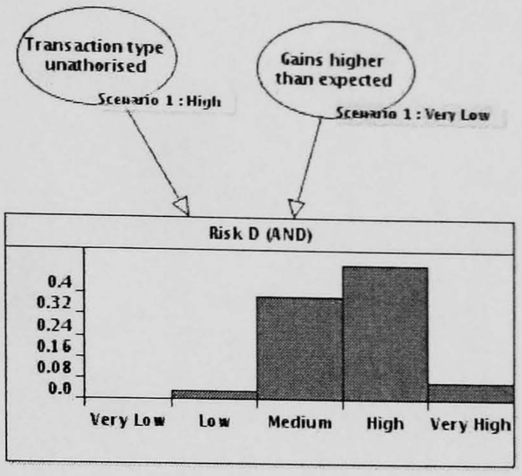


Figure (b). The probability of a "Transaction type unauthorised" is "Very High" and the evidence on the "Gains above expected" is "Very Low".

Figure 7.10: Example of an "AND" gate node. Compare figures (a) with (b). Despite adding the evidence on the node "Gains above expected" (see Figure (b)) the probability distribution for the child node does not change.

We can expect the level of risk shown in the KRIs to be related with the firm's management. Thus we have the link from the "Organisational factors" module in Figure 7.9. The quality of the BL's control becomes a reflection of the organisational factors and the supervisory duties at BL level. This quality tells us the probability of potential events being capture in that BL.

7.5.3 Organisational factors

As the Basel Committee [164] points out, one of the main areas of concern in the risk management process is the establishment of a strong culture of control. To measure the strength of these controls we must look at factors such as the firm's culture, the employee's performance and the quality of the support systems. We have decomposed these organisational factors into three modules, see Figure 7.11 in page 149:

- People
- Culture
- Systems

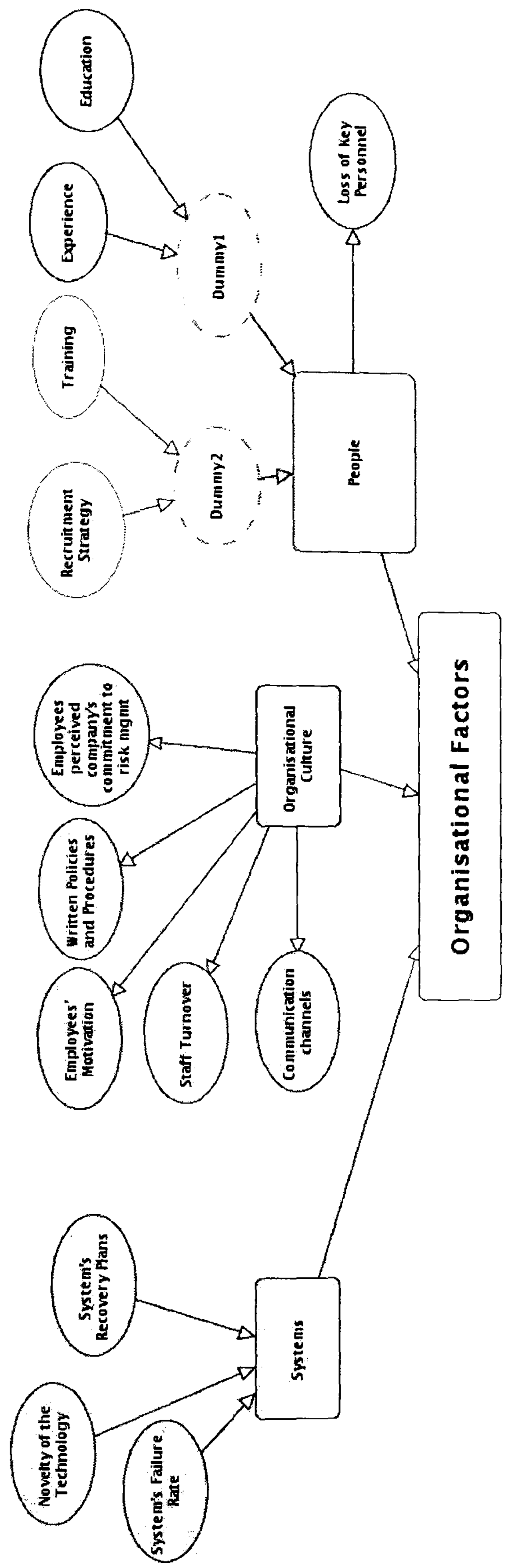


Figure 7.11: Organisational Factors module

7.5.3.1 People

This module captures the quality of the staff. This quality, among other things, depends on employees' capabilities. We defined those as a function of their level of "Education" and "Experience". The Basel Committee [106] adds another two shaping attributes: the firm's "Recruitment strategy" and staff "Training" as being the first step in assessing a firm's risk profile: and the loss of key personnel. As Figure 7.12 shows the resulting module has 4 input nodes and 2 output nodes, "People" and "Loss of Key Personnel". The latter node stands for the personnel whose role is supervising and mentoring other members of the staff. As such it can be used as an indicator of the staff quality. Notice that to reduce the size of the output node's NPT we have *divorced* the parent nodes into two groups.

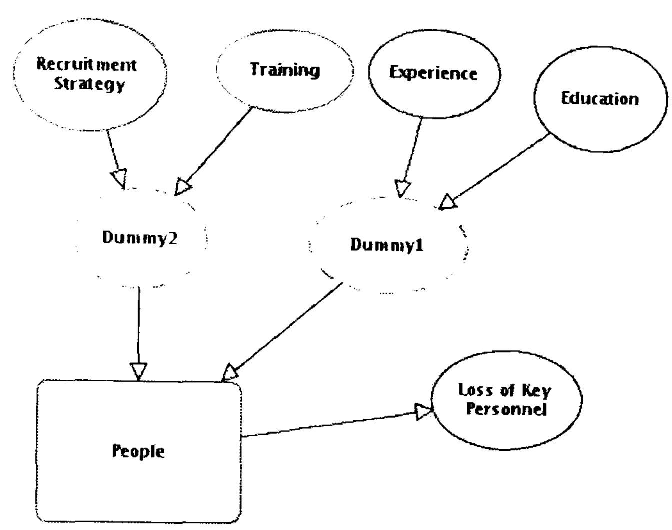


Figure 7.12: Sub-module People. The intermediary "Dummy" nodes to reduce the size of "People" node NPT.

7.5.3.2 Culture

This module is based on the assumption that culture is a hidden factor whose definition we ignore, that is, we cannot define directly what culture is. Thus, we need to look at the factors that "successful" companies exhibit, to define. in terms of performance, what makes a "good" or a "bad" culture. Only by looking at these shaping factors are we able to define culture. The factors shown in Figure 7.13 were drawn from our discussion on the role of the organisational culture in chapter 2, these are:

- Employees perceived commitment by the management towards risk strategies.
- Written policies and procedures.
- Employees' motivation.
- Staff turnover (this refers to company's employees in general not only "Key person-

nel".).

- Communication channels.

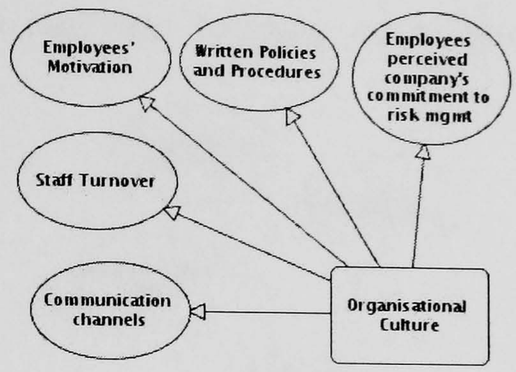


Figure 7.13: Sub-module Culture

Looking at the evidence available on these factors we can infer the culture that better explains these factors' evidence. For instance, if we observe that the "Communication channels" between employees and managers are "High", the "Staff Turnover" is "Medium", the "Employees Motivation" is "Medium", the written "Policies and Procedures" are "High" and the "Employees managers commitment" is "Medium" we would expect a culture, thus described, to be mainly towards "Medium", see Figure 7.14. Which means that the factors "Policies and Procedures" and "Communication channels" are not relevant enough to qualify the culture of a company.

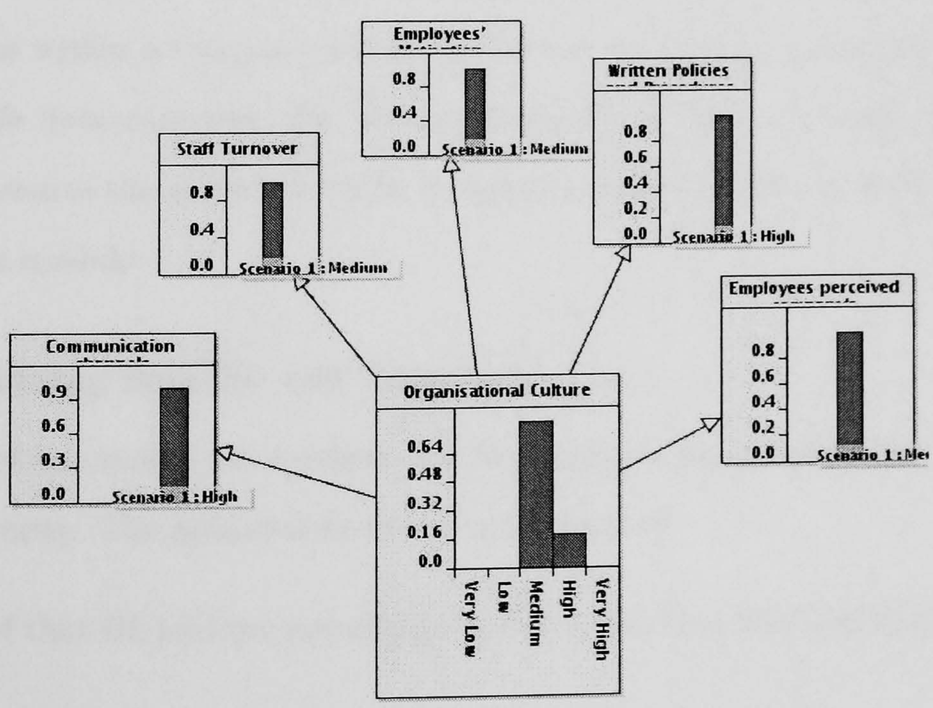


Figure 7.14: Culture's module with evidence entered

7.5.3.3 Systems

This module represent the expected performance of the company's systems support. This performance is affected by the impact of the "Novelty of the technology" (e.g. e-commerce), "System failure rate" and "System's recovery plans". See Figure 7.15.

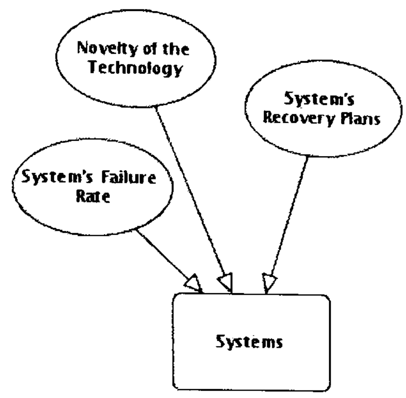


Figure 7.15: Sub-module Systems

7.5.4 Controls

This module represents the auditing controls, whether internal or external, of the financial institution. The auditing process depends, to a great extent, on the amount and quality of the losses recorded. For this reason, as we discussed in chapter 2, company's management must ensure a clear line of reporting between employees and managers. Thus, the causal links between the organisational factors module and this one. We can also expect that different teams within a company will accommodate the company's culture to suit their tasks [78]. We have captured this "decentralisation" in the node called "Sub-culture". This node measures the extent to which company's culture applies to them. Figure 7.16 illustrates this module.

7.5.5 Combining Severity and Frequency

In this part of the model we combine the frequency of the potential losses and their associated severity. The potential losses are a function of:

- the GI of that BL and the percentage of risk events that that slip through the BL's controls
- and from the quality of auditing controls whether internal or external to the firm.

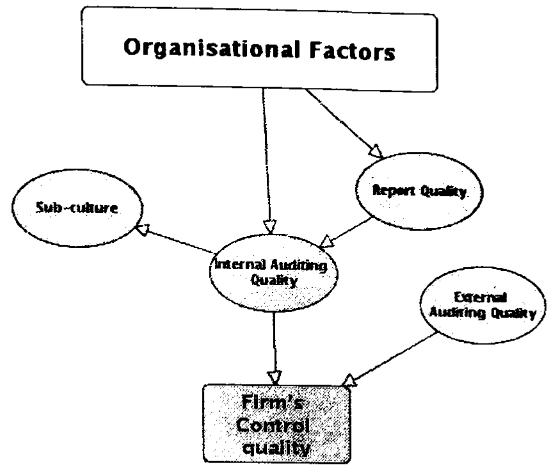


Figure 7.16: Module Controls

The result of combining the potential losses and the controls gives us the forecasted frequency of events. Aggregating this frequency with the severity per event will give us the total amount of predicted losses for this BL. Notice that we have conditioned the node "Severity" on the quality of the firm's management. It is reasonable to think that a firm with strong controls will be able to handle the severity of the losses better than another one that has not previewed those losses or established contingency plans to contain them, (see Figure 7.17). Figure 7.18 in page 154 shows the final OpRisk model for the "Trading and Sales BL".

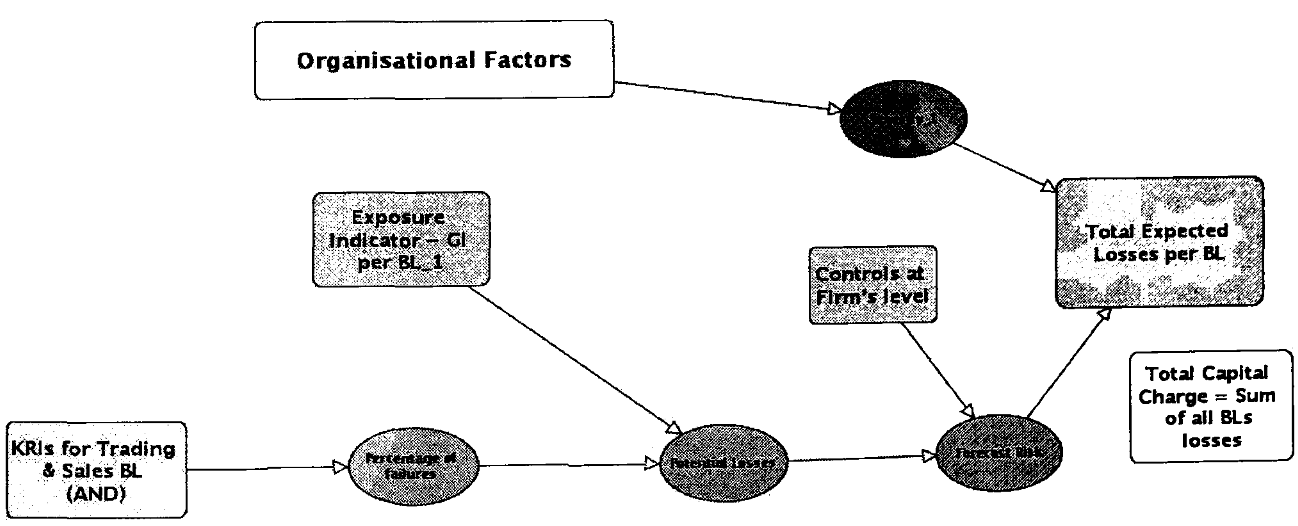


Figure 7.17: Joining the modules.

7.6 NPTs

The techniques used to build the NPTs of this model were explained in chapter 5. The nodes in this model are, by default, *ranked nodes* with 5 states ranging from "Very Low",

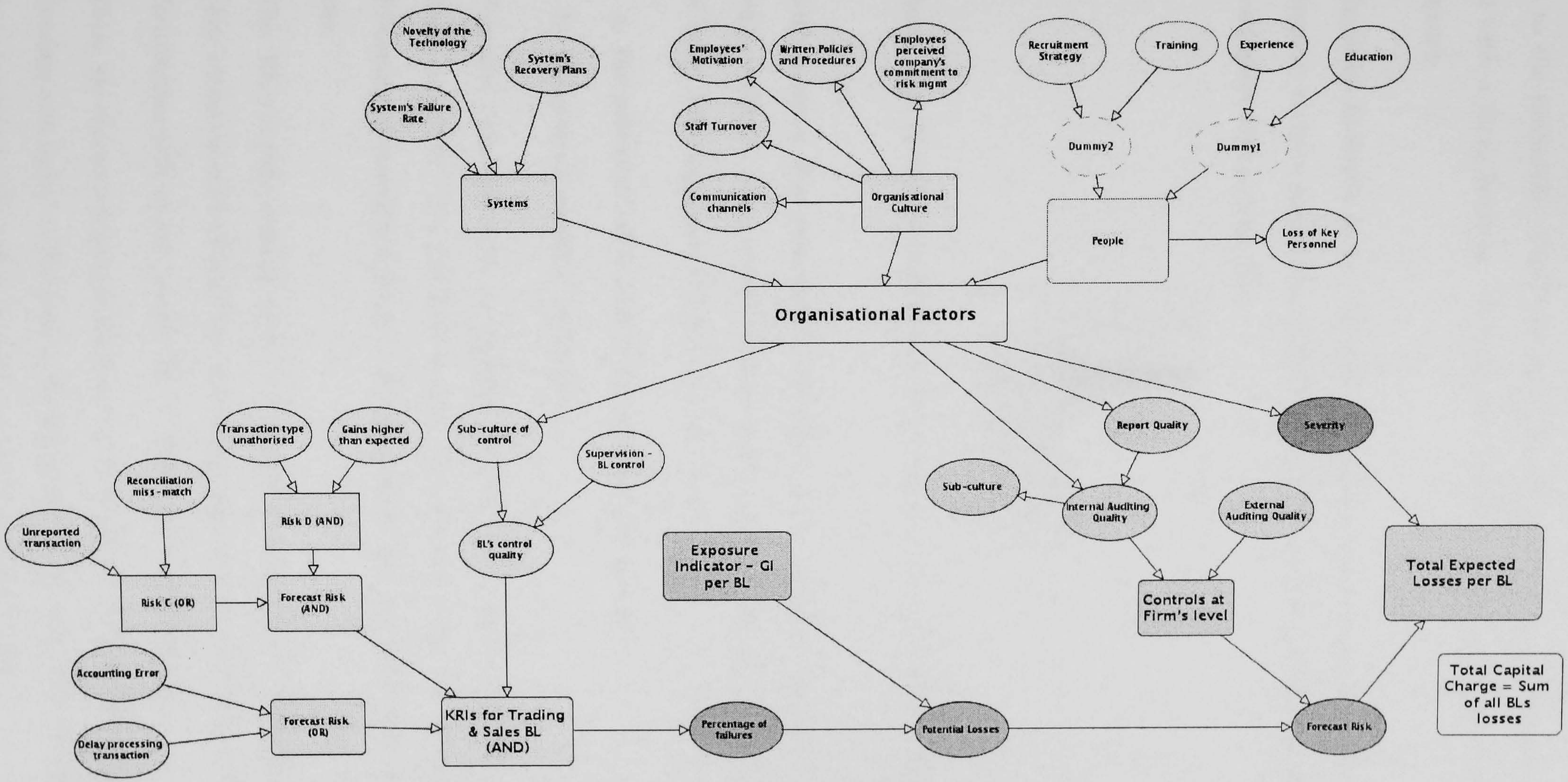


Figure 7.18: Final OpRisk model.

"Low", "Medium", "High" to "Very High", unless explicitly stated otherwise. The same applies to the probability distributions which are built using a TNormal distribution coupled with a *fuzzy function.* The following list provides an overview of the NPTs of this network:

1. Exposure Indicator node. For lack of a better indicator and to make calculations simpler, we have assumed a GI for this BL to be between 26.000 and 27.000 euros, see Figure 7.19 in page 155.

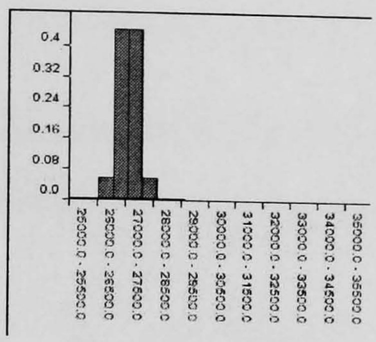


Figure 7.19: Normal distribution with a mean of around 26000 to 27000 euros.

2. KRIs module. The assessment of the KRIs can be based on the historical loss data, when available, or on a subjective estimation using Scorecards, or the combination of both. This module is divided into two parts, .

   a. the potential risk arising from business processes and

   b. the controls available at BL level.

   Potential risks may arise in conjunction with other risks, i.e. "AND", or either of them, i.e. "OR". For the former case, the expected value, i.e. the mean for the TNormal, is calculated using a *MAX* function and the latter case using a *MIN* one.

   The "BL's control quality" node is a weighted function of the strength of the company's internal controls and the supervision management at BL level. The resulting distribution tells us the probability of capturing a potential risk.

   Thus, we choose a binomial distribution to build the NPT for the node "KRIs for Trading and Sales" to forecast potential losses at BL level with the probability of success given by the "BL's control's quality", see Figure 7.20.

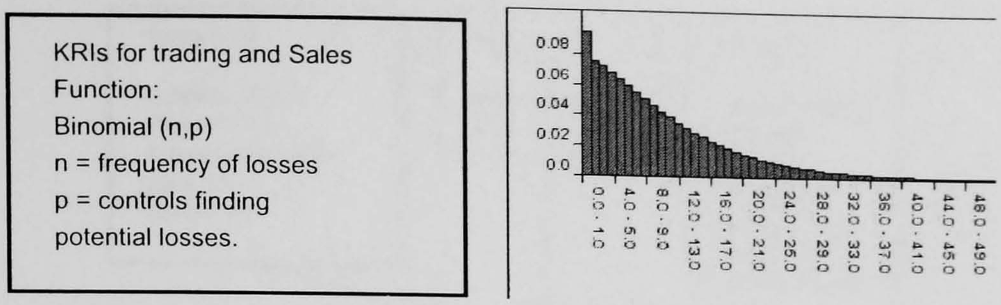


Figure 7.20: Binomial distribution for "KRIs for Trading and Sales".

3. Organisational factors module. The NPT for this module is captured in the node "Organisational Factors". This NPT is an average sum of the other 3 sub-modules:
   (a) People. This submodule is made up of 5 input nodes and 1 output node called "People". We have modelled the output node "People" as the weighted sum of its attributes.
   (b) Organisational Culture. The probability estimates usually come from either questionnaires as we did in our case study with NATS explained in chapter 6 or using a Scorecard approach. There is no predefine function to build the NPTs of this submodule other than the one that better fits the given information.
   (c) Systems. This submodule is made up of 3 input nodes and 1 output node called "Systems". The measurement of the input nodes "System failure rate" and "System's recovery plans" can be obtained using statistical metrics from past failures while the factor "Novelty of the technology" can be obtained through a Scorecard approach. The relevance of these factors towards the system's performance are captured using a weighted-sum function.
4. Controls module. This module is made up of 3 input nodes, an intermediary node labelled "Internal Auditing quality" and an output node labelled "Firms control quality". The NPT for the node "Firms control quality" is determined by a weighted sum of the auditing controls.
5. Potential risks node. This NPT is a function of the percentage of events that slip through BL's controls times the volume of the BL's business.
6. Forecast risk node. The NPT is built using a Binomial distribution whose probability is determined by the quality of the firm's controls and the potential event to

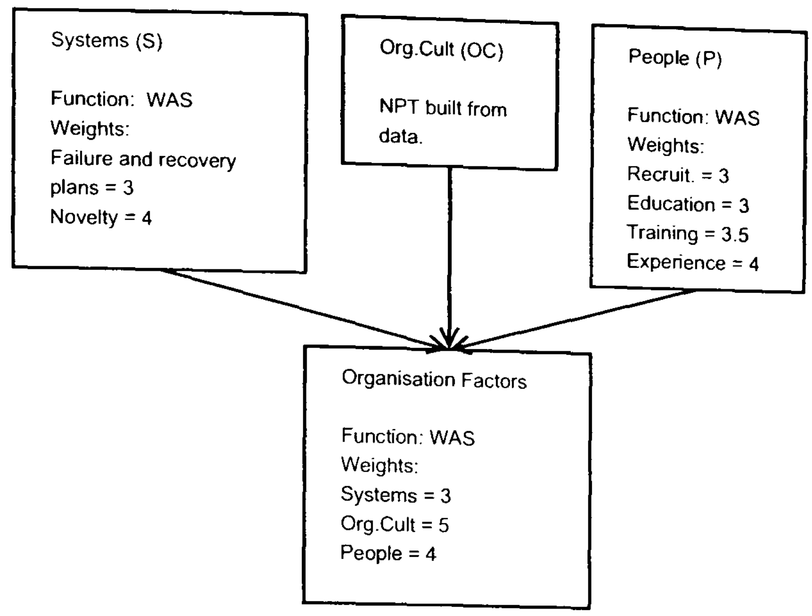


Figure 7.21: Fuzzy functions used to build the Organisational Factors module.

cause losses.

7. Severity node. This node is defined by a Chi-square distribution using the values from table 7.22 in page 158 as a guide.

8. Total expected losses node. The NPT for this node is the result of aggregating the distribution for the severity and forecast. The result obtained is the total losses for that BL.

7.7 Validation

In this section we carry out a similar validation process that we did in section 6.8. This process consists of identifying specific variables in the network, entering evidence and analysing the associated changes in the other variables in the network.

The loss events and loss amounts used to quantify the variables in the model are not indicative of the actual values in a real financial institution. As a guidance, we used the values obtained from the second qualitative impact study conducted in 2002 by the Committee [165], these are shown in tables 7.22 and 7.23.

- Scenario 1. The default settings for this model are (see Figure 7.24 in page 159)

  1. GI around 26.000 to 27.000 euros.

  2. The forecast individual loss events is between 1.100 and 1.300, see table 7.23.

	Internal Fraud	External Fraud	Employment Practices and Workplace Safety	Clients, Products and Business Services	Damage to Physical Assets	Business Disruption and System Failures	Execution, Delivery, and Process Management	Total Across Event Types For Each Business Line
Corporate Finance	3.293 0 13%	25,231 0 97%	6,114 0 23%	131,012 5.01%	18 0.00%		28.432 1.09%	194.100 7 43%
Trading and Sales	68,819 2 63%	828 0 03%	7,845 0.30%	89.054 3.41%	138 0.01%	6.237 0.24%	326.563 12.50%	499.461 19 11%
Retail Banking	115,578 4 42%	210.028 8 04%	54,600 2 09%	387,447 14 83%	61,178 2 34%	2,110 0.08%	198.620 7 61%	1.029.757 39 41%
Commercial Banking	78,889 3 02%	287.855 11.02%	3,662 0 14%	76,217 2.92%	14,033 0.54%	1,424 0.05%	136.659 5.23%	598.717 22 91%
Payment and Settlement	750 0.03%	5,447 0.21%	719 0.03%	1,144 0.04%	2,061 0.08%	2,705 0.10%	112.468 4.30%	125,295 4 79%
Agency and Custody Services	2,285 0 09%	281 0.01%	374 0 01%	7,635 0.29%	860 0 03%	1,718 0 07%	43.310 1 66%	56.443 2 16%
Asset Management	8,568 0 33%	603 0 02%	1,075 0 04%	8.978 0.34%		664 0.03%	34.641 1.33%	54.728 2 09%
Retail Brokerage	445 0.02%	596 0 02%	1.645 0.07%	17.485 0.67%	575 0.02%	6.471 0.25%	27.127 1.04%	54.545 2 09%
Total Across Business Lines	278,566 10.66%	530.868 20.32%	76,235 2 92%	71,8971 27.51%	78,880 3.02%	21.329 0.82%	908.219 34.76%	2.613,066 100.00%

Figure 7.22: Total Gross Loss Amounts by BL and ET thousands of euros

30 Banks Reporting Data

	Internal Fraud	External Fraud	Employment Practices and Workplace Safety	Clients, Products and Business Services	Damage to Physical Assets	Business Disruption and System Failures	Execution, Delivery, and Process Management	Total Across Event Types
Corporate Finance	4 0.0%	3 0.01%	16 0.06%	15 0.05%	8 0.03%	1 0.00%	33 0.12%	80 0.29%
Trading and Sales	16 0 06%	6 0.02%	37 0.14%	112 0.41%	10 0 04%	39 0.14%	1,114 4 07%	1.334 4 87%
Retail Banking	593 2.17%	7.796 28.49%	579 2.12%	1,273 4.65%	837 3.06%	570 2.08%	6.607 24.67%	18.457 67.43%
Commercial Banking	93 0 34%	1.180 4.31%	55 0.20%	66 0 24%	285 1.04%	474 1 73%	1.463 5 35%	3.618 13.21%
Payment and Settlement	22 0.08%	961 3.51%	9 0.03%	57 0.21%	40 0 15%	64 0.23%	752 2.75%	1.905 6.96%
Agency and Custody Services	6 0 02%	7 0 03%	12 0.04%	69 0 25%	17 0 06%	11 0.04%	356 1 30%	478 1 75%
Asset Management	4 0.01%	4 0.01%	21 0.08%	35 0.13%		6 0.02%	360 1.32%	430 1.57%
Retail Brokerage	7 0 03%	2 0.01%	12 0.04%	122 0.45%	28 0.10%	291 1.06%	609 2.22%	1.071 3.91%
Total Across Business Lines	745 2 72%	9,961 36.39%	741 2.71%	1,749 6.39%	1,225 4.46%	1,456 5.32%	11,494 41.99%	27,371 100 00%

Figure 7.23: Number of Individual Loss Events per BL and ET

30 Banks Reporting Data

3. According to table 7.22, the severity would be between 200 and 400 euros per loss event.

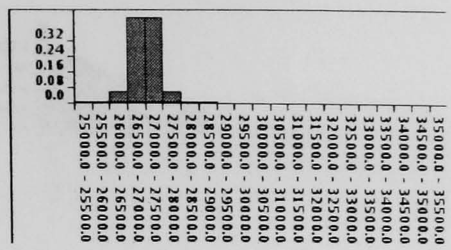
4. Therefore, the total losses are on average 400.000 euros.

Let us compare these default settings with other scenarios. To make this comparison easier we have drawn the probability distributions from the default scenario with bars while the distributions for the following scenarios are drawn using a line graph:

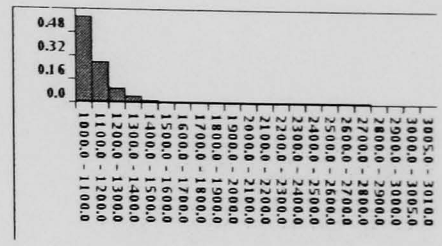
- Scenario 2. Operational Factors is set to "Low" (see Figure 7.25 in page 159).

  1. The expected risk events coming from the BL have increased approximately 30%. Observe how the tail of the distribution has now thickened.

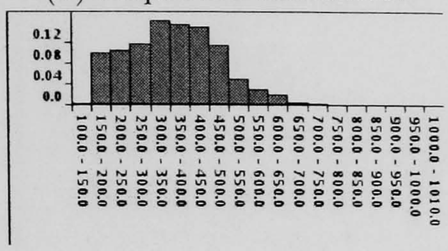
  2. this lack of management has also impact the severity which has risen on average



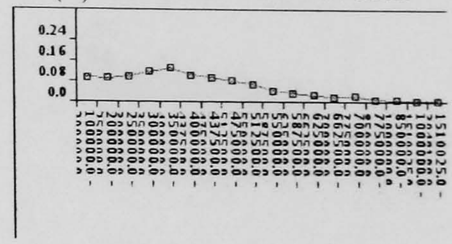
(1) Exposure Indicator.



(2) Forecast Loss events

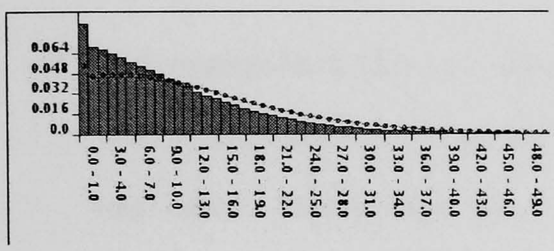


(3) Severity of of individual events.

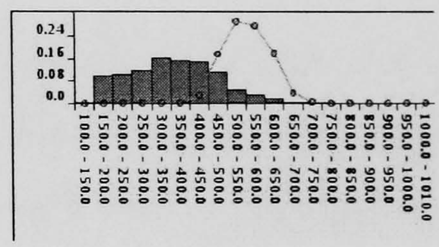


(4) Total loss of BL

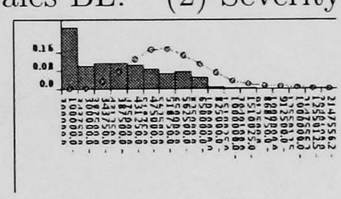
Figure 7.24: Default settings, previous to entering any evidence.



(1) KRIs for Trading and Sales BL.



(2) Severity of of individual loss events.



(3) Total loss of BL

Figure 7.25: Settings for scenario 2.

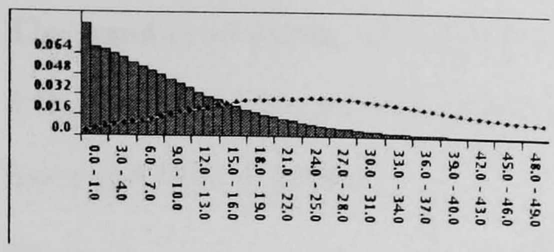
20% per loss event;

3. the predicted total loss has increased by 60% to be around 600.000 euros. But, even more worrying is the thickening of the distribution tail. This can give way to unexpected major losses.

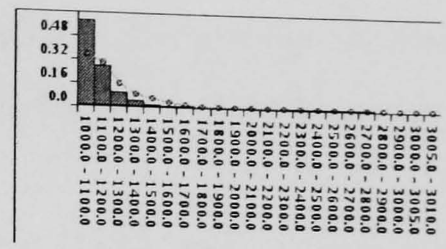
- Scenario 3. Controls at BL's level and Firm's level are set to "Low" (see Figure 7.26 in page 160).

  1. The loss events coming from the BL have more than doubled, from 10 to 25 expected events. Observe the tail of the distribution, the variance has also almost a doubled;

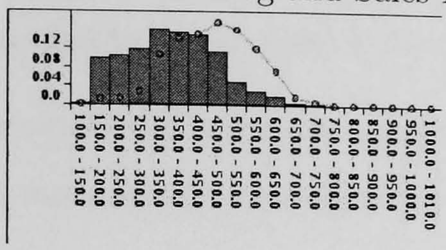
  2. The forecast losses increased slightly from 1100 events to 1200 but the variance widens by around 35%;



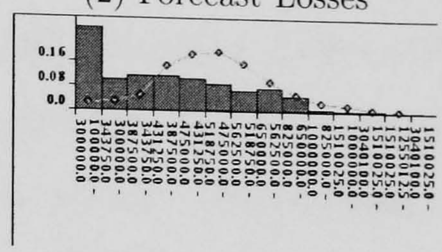
(1) KRIs for Trading and Sales BL.



(2) Forecast Losses



(3) Severity of of individual events.



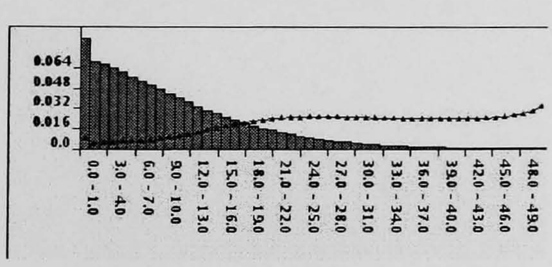
(4) Total loss of BL

Figure 7.26: Settings for scenario 3

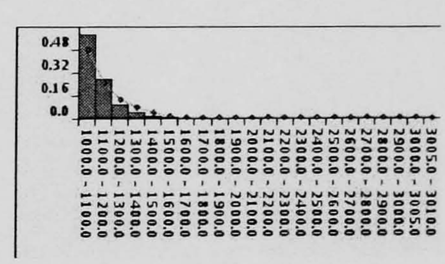
3. thus unexpected risks are more likely as the tail of the distribution gets thicker which, combined with an increase in the severity of approximately 30% per loss event , means that the expected losses are now around 534.060 euros for the BL.

- Scenario 4 has the following settings (see Figure 7.27 in page 160):

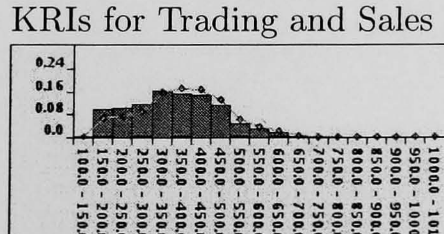
  1. Report quality is set to "Low";
  2. BL's supervision controls are "Medium", i.e. average;
  3. Transaction type is unauthorised "Very High";
  4. Reconciliation miss match is "Very High".



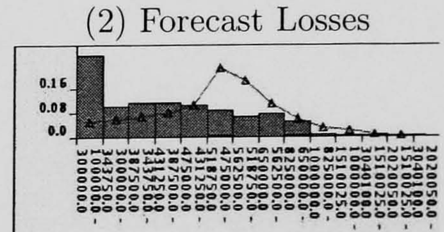
(1) KRIs for Trading and Sales BL.



(2) Forecast Losses



(3) Severity of of individual events.



(4) Total loss of BL

Figure 7.27: Settings for scenario 4

1. The number of events slipping through the BL's controls has now increased 3 times but more noticeable is the increase in the distribution variance which has more than doubled;

2. The same can be said about the forecast risk. The variance has also more than doubled, thus making the likelihood of unexpected events increase.

3. The severity, however, has not changed much. It has increased its variance and loss amount per event but not considerably. The reasoning behind this result is that the loss events come from the BL and the management of the severity is not directly link to the BL but to the firm itself. We can think of instances, e.g. AIB discussed, where a financial branch commits fraud and the headquarters do not realise those until is too late.

4. The total expected losses have increased on average to 40.000 euros but the variance remains approximately the same.

From these scenarios we observe how unexpected losses are explained from the expected ones. This relation is more noticeable as the tail of the distribution increases, thus making major losses more likely. This validation exercise has shown that the impact of the organisational factors in the firm's losses can be captured and measured using a BN model.

AMA			
Methods	Source of Information	Pros	Cons
Value-at-Risk	loss data	Well-known method. Good approach given the lack of a better one	It is meant for credit and market risk. Regulators want banks to move on from this approach.
Internal Measurement	GI per BL and ET and a given gamma γ factor	No need of historical loss data, but it can use it. More detailed than the Standard approach.	The efficiency of internal controls are not taken into account when calculating the charge. No sensitive to risk
Loss distribution	Internal and external data	Actuarial methods are often used to calculate the charge as a function of the frequency and severity of the given loss event	The need of historical loss data. The costs involved in building a database or buying external data.
Scorecards	Expert's judgment	It assesses bank's operations and activities against potential losses. Managers can take pro-active steps to reduce future losses. It provides a means to translate qualitative assessments into quantitative values that give a relative ranking of different types of OpRisks.	Relies on subjective estimation and as such it needs the employee's "willingness" to collaborate in the assessment. Banks following this approach will have to demonstrate a strong culture of internal controls in order to make their assessment valid.
Risk Indicators	It combines subjective and objective data	Their impact can be calculated using statistical regression. Managers can take pro-active steps to reduce future losses.	It does not say anything about the underlying process that makes these indicators change or interact.
Bayesian Networks	It combines subjective and objective data.	It describes the business unit process. It can complement historical data with expert's experience and knowledge. No need to invest on large databases. Adapts to fast changing environment	J. King [119] argues that BN is a good tool to capture expected losses but not the unexpected ones. The argument used is that causal reasoning cannot explain the latter.

Table 7.8: AMA approaches

Part IV

Summary and Conclusions

Chapter 8

Summary and Conclusions

The aim of this thesis is to show how large NPTs can be built using qualitative knowledge in a consistent and economical manner. Large NPTs are the consequence of complex BN models. These models are needed to capture the many interactions among the risk factors in the safety critical industries and financial institutions.

This thesis takes the view that risk factors, forming a causal chain, shape the organisation's risk profile. These factors include the impact of the organisational culture, the quality of the staff and the infrastructure of the organisation's supporting systems, what we have termed as *organisational factors*.

A *deep* understanding of these risks enables us to interpret them as causally related; measuring the probability of these *latent failures* along the causal chain gives us the opportunity to take steps to avoid, or at least to reduce the impact of *active failures* on the organisation [190], e.g. ENRON [194], or on society as a whole, e.g. Bhopal chemical plant [145].

One of the most important challenges in modeling real world BNs for risk assessment is that of producing large NPTs.

This thesis explains a method and implemented a semi-automated computer program to build them (see Appendix A) which is now part of the Agena Risk tool [134]. We have found that in the case of the NATS project this approach and the program made the difference between being and not being able to build realistic a BN model within the constraints of time and costs.

8.1 The Ranked Node Approach

As we discussed at the beginning of this thesis, one of the challenges in building BN models to solve real-world risk assessment problems is that of constructing large NPTs. To produce large NPTs we need to involve domain experts (who do not necessarily have knowledge on probability theory) whose time is sparse and costly.

For these reasons, we need a method to construct NPTs using the minimal amount of expert elicitation, recognizing that it is rarely cost effective or feasible to elicit complete sets of probability values.

In the experience of RADAR's group building BNs, we know that most of the time this relationships can be expressed as a weighted average function. Many so-called "self-assessment" or "scorecard" [21] systems are based around little more than the weighted averages of attribute hierarchies.

This weighted average function can be interpreted as a linear relationship between parent and child nodes. Simple linear models have shown to provide good forecasts as we discussed on Chapters 4 and 5. Furthermore, in some cases, complex statistical functions have shown to add little benefit to justify the time and cost put into it [79]. In the case of the NATS project we could not see techniques that elicit degrees of freedom or matrix of covariates as possible candidate methods [183] given the lack of statistical knowledge of the participants.

Furthermore, when the assumption of linearity cannot be maintained we can use the WMin or WMax functions; for instance, a parent node that outweighs the impact of other parent nodes, this is explained in Chapter 5.

Although this approach may seem coarse we must take into account that we are talking about uncertain domains and that scenarios are elicited, as Parsons and Saffiotti [173] comment

> ...in the form "if we observe A then it is more likely that B is the case", we may be more interested in knowing the way in which the values change rather than in the values themselves.

So the aim of this type of model is as Jensen [111] observes to find the patterns that can be extrapolated to future events given our current knowledge.

> ...correct findings originating from a coherent case covered by the model should conform to certain expected patterns ...

The question would then be whether these patterns are informative enough to the expert. Van der Gaag et al's [219] research answers this question when they comment comment regarding their findings modeling oesophagus cancer:

> ...we would like to note that the data collection used is known to be biased, to contain inconsistencies and to be incompleted in a non-random way ... the percentage of correct predictions ... approached 70%. Given that the probabilities used are rough, initial assessments and that the patient data definitely require large clearing out. the results from initial evaluation are quite encouraging.

It is for these reasons that we advocated a general, flexible approach that can accommodate different types of relationships, rather than a bespoke distribution whose development would be lengthy and costly. A simple approach that captures the expert's opinion and not their knowledge in statistics.

As physicist S. Hawking [94] notes, writing about planet Mercury's motion, that although Einstein's theory of relativity is more accurate predicting this motion than Newton's law of gravity, Newton's law is the one used for all practical purposes given that the differences between both predictions are small and

> Newton's theory also has the great advantage that is much simpler to work with than Einstein's!

Note that using BNs the model's uncertainty is reduced as the system gathers data given that, as we discussed in Chapter 3, the expert's information is conditioned on the evidence at hand. As this evidence grows, any disagreement between expert's opinion and factual information is adjusted.

This does not, however, mean to say that in all cases using rank nodes guarantees better results than using other approaches. Theses results are based on these constraints:

1. The qualitative information as the main source of information;

2. The domain experts we worked with, who did not have statistical background, were able to build and tailor large models that effectively captured their beliefs well;

3. The time and budget allowance. Using rank nodes the elicitation burden is much reduced by simply eliciting a small number of parameters from the experts.

8.2 Lessons Learnt

These lessons are mainly drawn from our experiences developing the NATS project. These experiences helped to put into context the theory and the practicalities of building a BN model; to consolidate the theory and also to question it. It consolidated the understanding of the elicitation process; knowing what to expect and how to manage it. It also questioned the methods reviewed in Chapter 4 to a great extent.

The early stages of building the model were daunting because of the steep learning curve for all participants. In the case of the NATS project, most of the participants were alien to the concept of conditional probability; they were more acquainted with risk management.

The interpretation of BN as a causal model was helpful to establish a common language. However, some of the participants were reluctant to accept that node's relationships were causal. Their objection was that causality could not be ascertained even though they could conceive a relation as causal; for some causality owed more to the field of philosophy than it does to the domain of statistics. This objection changed when they understood how the BN model explained its outcomes and which causal factors were more important, e.g. pilot's skills, Culture.

The model structure helped to understand this idea of causality. In a barrier model we can observe how the risks progress from one barrier to the other this progression is caused by a barrier not being able to contain the risk thus having an effect on the next barrier.

It was also important to explain that there are many potential models that can explain a relationship. A relationship that is explained in probabilistic terms and whose outcome is interpreted as a *best fit* of the current model.

With the conceptual model came the problem of overlapping concepts, e.g. How do we model the effect of weather on an air incident? is that an independent factor or does it have an impact on other factors? Problems of overlapping concepts such as the weather were interpreted as contextual probability evenly distributed among the factors (see section 3.6 in 35).

When modeling the impact of culture in the organisation we had to challenge the idea that organisational culture could not be measured. This perception changed when the data from the research of our colleagues in this project, SRU, was put into context. Then, they acknowledged the importance of having this factor measured given its impact on the overall air-traffic strategy.

Another issue of discussion was the level of detail. Detailed definitions meant that the expert is able to discriminate a variable's impact on the model, can we differentiate the TCAS's input in an accident from that of the air-crew? The BN assumption of variable independence was a valuable tool on this case, e.g TCAS's advise is independent of pilot's skills, a pilot may or may not follow its advise. However, we could not say the same about the pilot's skill and the aircraft's crew. Both are understood as part of the same definition.

Finding a suitable model meant to re-make the network a few times before we agree on a final version. It was difficult to decide when to stop discussing the topology and concentrate on the probabilities. This decision had to take into account the time/cost of the model.

It is certainly important to be in control of the modelling process and explain the reasons for modelling certain relationship, e.g. constraint the number of parent nodes to reduce NPT's size or omit some factors, e.g. weather. The aim was to restrict the number of variables to those the expert could explain using a feasible scenario and whose outcome could be tractable using the SSE data (although it was not easy to directly map incident data into the model because some incidents were incorrectly or ambiguously recorded).

The constraints of time, budget and statistical knowledge implied the need for a flexible, general approach to capture expert knowledge. This is where the idea of ranked nodes originated from.

I implemented a computer program using the ranked nodes approach (see Appendix A). This made possible to build large NPTS, modify node's weights and Normal distribution deviation and observe the overall impact. Following a sensitivity analysis [40] approach we compared different inputs (see section 6.8 in page 114). The validation was carried out comparing the averages of the model outcomes against SSE data.

One of the findings by the NATS experts was that STCA was actually more relevant in preventing a mid-air accident that previously thought. To such an extent that the HR director proposed the use of the BN model to support the case for financing the installation of the STCA system in other airports.

8.3 Future Work

- A second phase of this project should include an evaluation following SSM criteria in terms of:
  - Efficacy. The model's advice has shown to have improved safety.
  - Efficiency. In economic terms, the ratio between the benefits of an improved safety record, derived from model's advice, and the cost to implement it.
  - Effective. Is the BN model's advice a good solution in the long run? Is it worth to invest on this solution?
- If the ranked approach, based on qualitative information, is to be used to monitor changing performance and risks over time it will be necessary to establish procedures for capture this information on a more or less regular basis. For instance, mechanisms to obtain subjectives assessment of the organisational culture, may need the introduction of routine reporting of some figures (e.g. SAQ questionnaire). Ideally to be able to obtain such data on-line and to introduce on-line monitoring.
- The NATS model was, by necessity, the product of the views of a reduced number of the Air-Traffic com
- On this latter point, we can see ontologies helping to unifying the criteria regarding different domains/concepts across an organisation. Future research should make use of tools like Protégé to develop BN models and see the applicability of our approach on different domains.
- To automate obtaining the weights β of expert's elicitated estimation. We can

interpret the scenarios elicitated as a system of equations, solving the equation system gives us the weighs or coefficients of each variable. parent nodes.

Eliciting different scenarios provides a system of equations from which we can obtain the weights or coefficients β.

For example. on Figure 8.1 we observe that the parent nodes "X: Experience" and "Z: Skills" condition the "Y: Managerial Quality", hereafter we refer to these variables as X, Z and Y respectively. Let us assume that all of these variables have these states: $\{verylow, low, medium, high, veryhigh\}$.

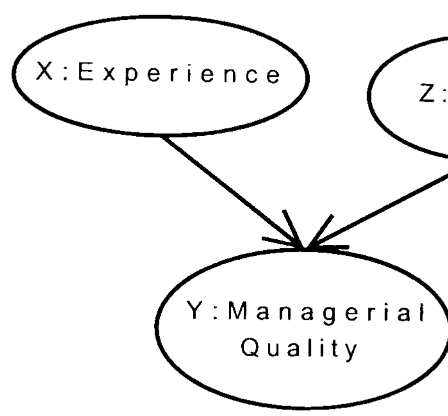


Figure 8.1: Managerial Quality

Note that following the ranked-node approach, explained in Chapter 5, these states equate to ordinal values from 1 to 5, e.g. very low = 1, low = 2. and so forth, so that we can effectively treat X, Y and Z as random variables in their full right.

A typical scenario to be elicited is for instance: given that the managers have no experience and have not got skills what is the expected managerial quality? i.e. "If Experience = low and Skills = low, what is the value of Quality?". If the answer to such questions is elicited as a drawing, a plausible set of data will look like Table 8.1, that also shows the mapping from the elicited drawings to their corresponding numerical values.

In Table 8.1 we also observe that, if we assume a multivariate linear relationship, eliciting different scenarios provides a system of equations from which the values of the regression coefficients β_x, β_z can be obtained. These quantities act as "weights" showing the strength of the relationship between the parents and the child.

In the case of this example, we observe that Experience is more important when

x_i	z_i	$= y_i$
5	5	5
1	5	2.8
5	1	3.8
3	3	3.5
1	1	1

$$\begin{cases} 5\beta_x + 5\beta_z = 5 \\ 1\beta_x + 5\beta_z = 2.8 \\ 5\beta_x + 1\beta_z = 3.8 \\ 3\beta_x + 3\beta_z = 3.5 \\ 1\beta_x + 1\beta_z = 1 \end{cases}$$

Solving the Multivariate linear regression, we obtain
E[y] = 0.625x + 0.375z + 0.22

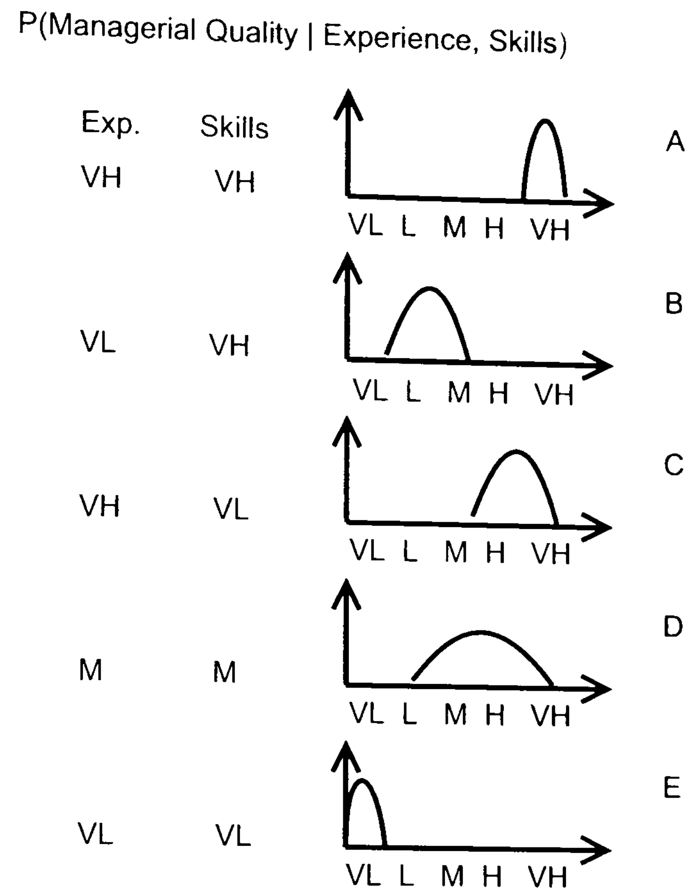


Table 8.1: Elicited drawings and values with their associated equation system.

considering manager's quality than is skill, as the ratio of the coefficients shows: $\beta_x/\beta_z = 0.625/0.375 \sim 1.6$.

- Using higher degree polynomials to represent non-linear relations. If we think that the current observations are going to remain constant in the future, e.g. domain of engineering, then a higher degree polynomial can provide a better fit. Otherwise, using a straight line can provide better results as it can be extrapolated to future events with greater success, although this may mean loosing accuracy on present observations. Figure 8.2 illustrates this idea.

Part of this work has been now incorporated into the Agena Risk tool [134]. This tool is the result of the continuous research carry out by the RADAR group at Queen Mary University.

This thesis has shown with some degree of success how expert's opinion together with BNs can be used to forecast risks in the safety critical industry and in the financial institutions.

It has produced promising results which, in view of the author, have warranted

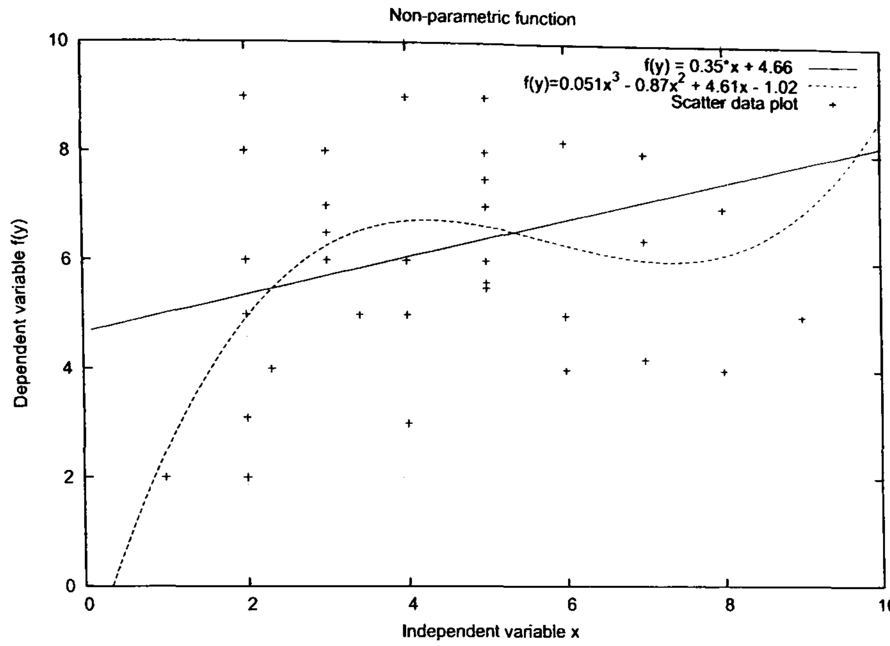


Figure 8.2: It depicts the situation where a data set can be modelled using either a straight line or using higher degree polynomial.

the need to carry out further research on the use of BN to assess risks. In particular, the author hopes to encourage decision makers to consider the use of BNs to handle organisational risks.

Appendix A. MiniHugin

Using MiniHugin

Part of this thesis was to implement a computer program to elicit and build NPTs putting into practice the ranked nodes approach. This approach is explained in detail in chapter 5.

Current BN tools like Hugin [13], Genie and Smile [65], Elvira [215] or Bayesian network tools in java [154] have not got the functionality required to develop the NPTs for the NATS project. We needed a tool that could model NPTs by domain experts without statistical knowledge using a few cues. These tools focus on calculating the model's output rather than to ease the elicitation of probability values.

The first prototype implemented the weight-average-sum function and used the odds function to calculate the deviation. This prototype was instrumental to realise a problem with the theory behind the odds function. This function failed to provide the expected resuls in the case of inter-causal reasoning.

The second prototype used a Normal distribution instead of the odds function to capture the deviation. Although the Normal distribution was more difficult to implement we were more confident that it would produce the expected outcomes given the supporting research in similar areas. These areas refered to the use of Normal distributions in domains were data is available.

Builiding this NPT using MiniHugin requires only to go through the following simple steps (see Figure 8.3):

1. Select a child node and define the NPT using an expression. This expression can be Average, Min or Max. The Normal distribution is selected by default.

2. Enter the elicitated values for the parents (and their weights if the expert is certain about strength of the relationship)

3. Enter the expected values for the child node given its parents and the "certainty" or variance value.

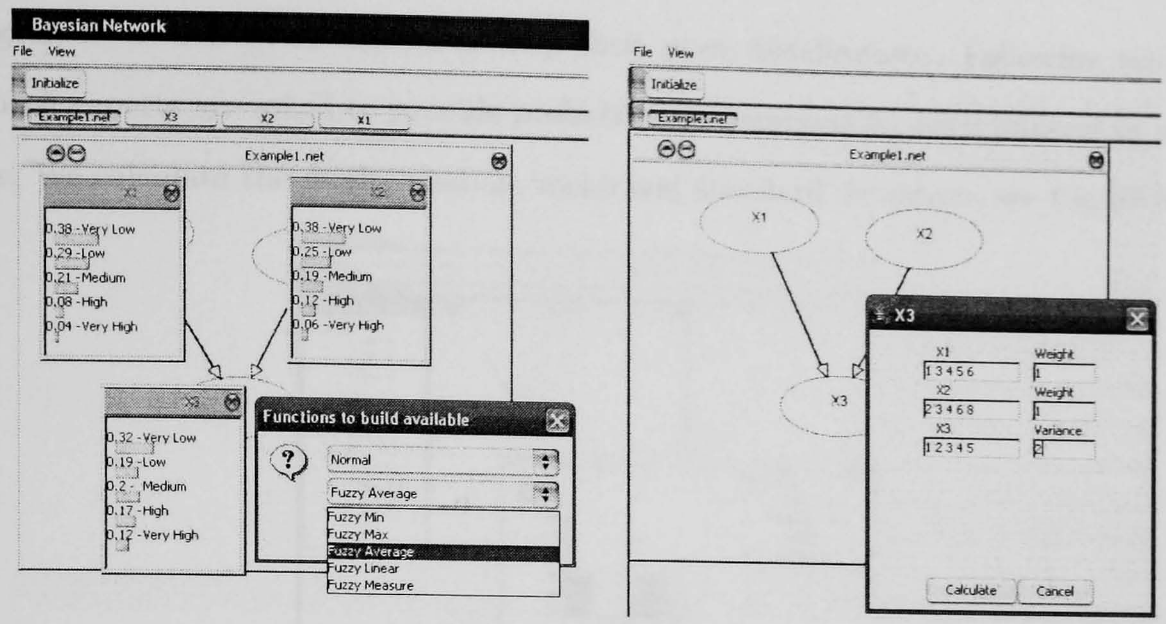


Figure 8.3: MiniHugin

As part of this tool, we also implemented some of the eliciting techniques reviewed in chapter 4, in particular from R. Winkler [127] research using drawings and intervals to elicit expert's knowledge, see chapter 4. These techniques are now explained:

Normal Distribution. It helps eliciting the Normal's mean and standard deviation, see Figure8.4. Using a slide bar the expert can select the mean and the deviation. These inputs are used to calculate the probability table for a given node.

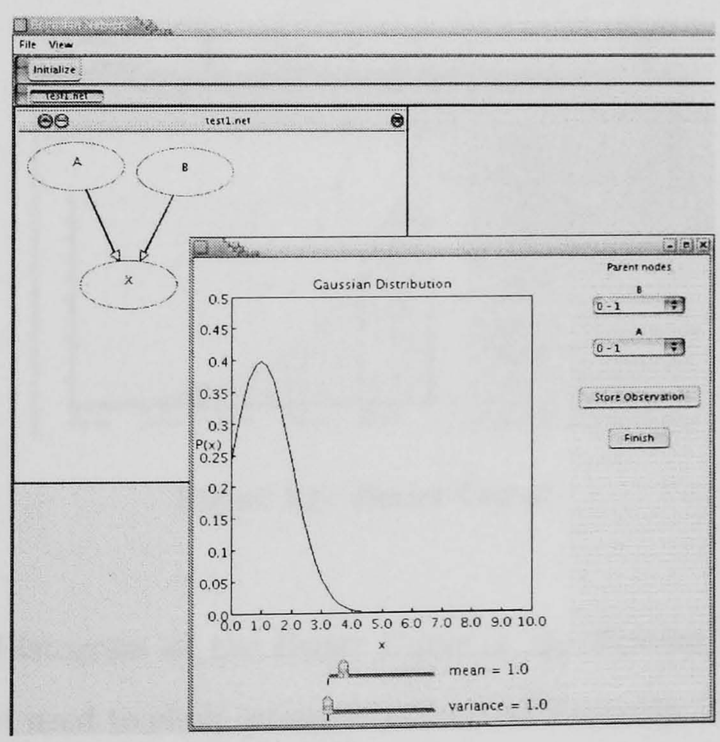


Figure 8.4: Eliciting a Nomal distribution.

Histograms. It use an histogram to help elicit prior distributions. Following the PDF approach experts are asked to provide probability estimations for each interval or node's states. We calculate the mode, median, mean and standard deviation, see Figure 8.5.

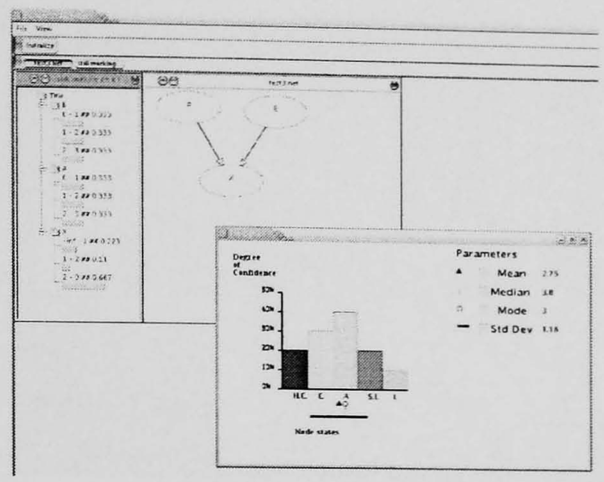


Figure 8.5: Histogram

Bezier Curve. Another tool was built to allow experts to draw the shape of a distribution. Using the *control points* experts can draw different shapes, see Figure 8.6. The distribution's area is calculated and normalised. The number of intervals in this tool vary from two, three to five. If possible, we approximate the distribution to a Binomial function.

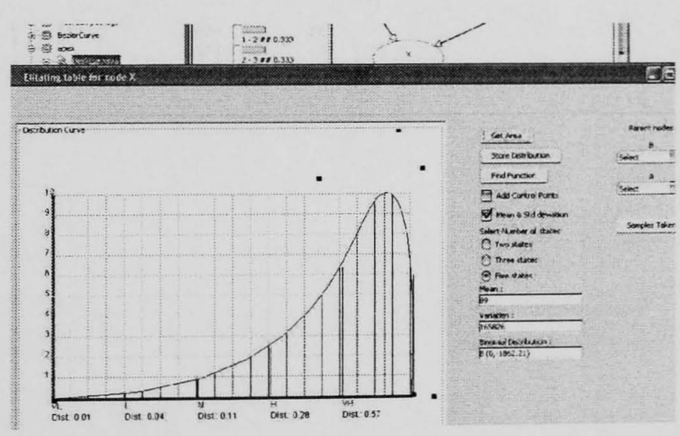


Figure 8.6: Bezier Curve.

Note that the Histogram or the Bezier Curve or the eliciting tool using a Normal distribution were not used to elicit estimates from NATS experts. They are the product of the lessons learnt during the model development and as a result of the thesis research. So, I cannot report how effective or otherwise these tools are.

These tools are now part of the AgenaRisk tool [134] which has its own algorithms to calculate the model's NPTs. Agena Risk has super-seeded this prototype providing a wider selection of methods to compute NPTs.

MiniHugin tool uses the Hugin [13] software library to calculate the network probabilities.

Appendix B. Safety Attitude Questionnaire (SAQ)

These are the collection of questions used by the SRA team to evaluate a company's safety culture.

Questions. Factor 1
I feel satisfied with the safety information I get
I am happy with the existing safety precautions for particularly hazardous work
I feel satisfied with the attention given to safety in any training I have had
I am happy with the safety equipment specified for my job
Generally I am happy with the safety in my asset area
I know the results of safety inspections to do with my job
The people I work with are satisfied with the attention given to safety in any training they have had
The people I work with are satisfied with the information they get about safe working
If changes are made to the procedures for my job I know about them
I feel I could tell my boss if I had worries about safety

Table 8.2: Questions related to "Personal evaluation of the safety system" factor

Questions. Factor 2
My safety representatives know about most aspects of safe working in their area
My safety representatives know about safety procedures
Safety committee meetings
If the safety representatives in my asset area/plant/department saw any one breaking safety rules they would do something about it

Table 8.3: Questions related to "Safety representative's perceived (knowledge of and) involvement in the safety system" factor

Questions. Factor 4
Management in the company encourage me to go to meetings which involve discussions about safety
I am encouraged by my supervisors to go to meetings which involve discussions about safety
Whenever there are meetings which involve safety discussions I go to them
The people I work with go to meetings which involve safety discussions
I am satisfied with the safety content of meetings we have
The people I work with are satisfied with the input they have at meetings where safety is discussed
My workmates know what has been discussed at safety committee meetings

Table 8.4: Questions related to "Workforce's perceived evaluation and involvement in safety meetings" factor

Questions. Factor 5
My supervisors know what safe working procedures people should be following
My supervisors know what safety equipment people in my asset area/plant/department should use
My supervisors know what has been discussed in safety committee meetings
The supervisors in my asset area know how much safety training people in my area have had
The management in the company know what safe working procedures people should be following
The managers in the company know what is discussed in safety committee meetings

Table 8.5: Questions related to "Management's perceived involvement (knowledge?) in the safety system" factor

Questions. Factor 6
To get their jobs done people in this asset area do not always follow safety procedures
The managers in the company think that following all safe working procedures gets in the way of production
I do my job in the same way as my workmates even when it means not always following safe working procedures
The management in the company puts productivity before safe working
My supervisors would turn a blind eye is safety rules were broken
The managers in the company do little to ensure that I follow safety procedures
I take short cuts to get my job done
I know of short cuts that would help me get my job done
If the people in my asset area notice a safety hazard they tend not to report it
My workmates do not know what they should do to ensure they are working safely

Table 8.6: Questions related to "Unsafe working practices" factor

Questions. Factor 7
My safety representatives are satisfied with the involvement they have in developing our safety training
My safety representatives are happy with the backing they get from management
My safety representatives are satisfied with their involvement in safety inspection
My safety representatives are satisfied with the authority they have to act in safety matters

Table 8.7: Questions related to "Safety representative's perceived evaluation of the safety system" factor

Questions. Factor 8
My workmates would expect me to support them if they had a complaint about safety
If I had a complaint about safety my workmates would support me
I encourage people in my asset area to work safely
The people I work with encourage me to work safely

Table 8.8: Questions related to "Workforce's (perceived safety encouragement and) support" factor

Questions. Factor 9
The people I work with check any safety equipment they might use before starting work
Generally my workmates keep the area they work in tidy
My workmates are satisfied with the safety procedures in general
The people I work with know what safety training is needed for their jobs
The people I work with understand the reasons for the safe working procedures they are supposed to follow

Table 8.9: Questions related to "Co-worker's perceived involvement and evaluation of the safety system" factor

Questions. Factor 10
My managers are satisfied with the safety procedures generally
My managers are satisfied with the results of safety inspections
My supervisors are generally satisfied with safety in my asset area
My supervisors are satisfied with the safety training given to their work group

Table 8.10: Questions related to "Management's perceived evaluation of the safety system" factor

Questions. Factor 11
My supervisors encourage me to report any safety problems I might notice
My supervisors talk to me about safe working procedures
The management in the company encourage me to let them know of any worries I have about the safety of my job
The managers in the company talk to me personally about safe working
The managers in the company encourage me to do housekeeping directly related to my job

Table 8.11: Questions related to "Participative communication" factor

Questions. Factor 12
Before I start work I check the safety equipment I might need
I know the written safe working procedures for my job
Generally I keep the area I work in tidy

Table 8.12: Questions related to "Personal involvement in the safety system (safe working practice?)" factor

Appendix C. Contributions

The paper *Using Ranked nodes to model qualitative judgement in Bayesian Networks* [148] written by M. Neil, N. Fenton and Jose Galan is the product of this thesis.

The idea of using the Weighted Average Sum, Weighted Min and Weighted Max functions to build NPTs corresponds to Norman Fenton and Martin Neil. I implemented the tool that contributed to confirm and validate their use to building large probability tables.

The first implementation of this tool used the odds function to capture the deviation of these functions results. However, as we explained in Chapter 5 we had asymmetry problems during back-propagation. A Normal distribution was used instead. Although using a Normal distribution was a chanllenge, from the point of view of programming, it proved to be the right choice, as the NATS project have shown.

I learnt about use of a Normal distribution from the book by Rusell and Norvig [197], in this case, I only applied their ideas in a framework where only qualitative information was available.

This tool made possible building all the NPTs of the NATS project. The outcomes were validated in the first instance by the SSE data. Later, those results, were calibrated by the experts at NATS and approved by the SRU. M. Neil and I run sensitivity analisys with the NATS's experts. This tool uses the Hugin machine library to make the calculations [13].

I was responsible for the NATS validation. This validation was done in terms of averages, e.g. at each barrier and at any given time a number of breaches are expected. The aim was to match the average of those breaches against the model's output.

The development of the Culture sub-network was my own contribution. Our colleagues at the SRU in Liverpool provided me with the data from the SAQ questionnaire in excel format. It took me a great amount of time, I had to go back and forward until both the SRU and NATS agreed with the results. Martin Neil overlook the whole process.

The Operational Risk chapter is the result of my own contribution. This was, at first, the aim of my thesis. I studied to a great depth the work of Basel Committee regulating the financial institution's operational risk. However, we could not find a financial instititution that wanted to collaborate with us on the study of the impact of culture in the firm's overall risks.

Note: During this PhD. I have benefited from the support and advise of my tutors Prof. Martin Neil and Prof. Norman Fenton. All the contributions I have personally made to this thesis give equal credit to my tutors who made them possible. Thank you.

Bibliography

[1] Daneshkhah A. Psychological aspects influencing elicitation of subjective probability. BEEP working paper, 2004.

[2] O'Hagan A. Eliciting experts beliefs in substantial practical applications. *The Statistician*, 47:22–35, 1998.

[3] O'Hagan A. and Oakley J.E. Probability is perfect, but we can't elicit it perfectly. *Department of Probability and Statistic. University of Sheffield.*, 2002.

[4] A. G. Castro, P. Rocca-Serra, R. Stevens, C Taylor, K. Nashar, M. A. Ragan and SA Sansone. The use of concept maps during knowledge elicitation in ontology development processes - the nutrigenomics use case. *BMC Bioinformatics*, 7:267, 2006. http://www.biomedcentral.com/1471-2105/7/267.

[5] P. Slovic A. Tversky and D. Kaheman. *Judgement under uncertainty heuristics and biases*, chapter 1. Judgement under uncertainty heuristics and biases. Cambridge University Press, 1982.

[6] Philips Innovative Applications Agena Ltd., IAI, Philips Digital Systems Laboratories and QinetiQ. Modist. Technical report, Commission of the European Communities under the Information Society Technologies programme (Project IST-2000-28749), Sept 2001 - Feb 2004.

[7] C Alexander. Bayesian methods for measuring operational risk. Technical report, The business school for financial markets, 2000. University of Reading.

[8] C. Alexander. *Operational Risk: Regulation, Analysis, and Management*, chapter 7. Statistical models of operational loss. Prentice Hall, 2003.

[9] V. Alexiev. *Information Integration with Ontologies: Experiences from an industrial showcase*. Chichester: Wiley and Sons ltd, 2005.

[10] Allied Irish Banks Plc. *Report to the Board of Directors of Allied Irish Banks Plc.* 2002.

[11] AmeliaFinancial. Operational risk: The last of the risk frontiers.. 1988.

[12] S. Andreassen, M. Woldbye, B. Falck and K. Andersen. Munin - a causal probabilistic network for interpretation of electromyographic findings. In *Proceedings of the Tenth International Joint Conference of AI*, volume 10. Morgan Kaufmann, 1987.

[13] Hugin A/S. Hugin software. www.hugin.com, 2005.

[14] Al awadhi S. A. and Garthwaite P. H. An elicitation method for multivariate normal distributions. *Communications in statistics. Theory and methods*, 27:1123–42, 1998.

[15] P Ayton and I Fischer. The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory and Cognition*, 32:1369–78, 2004.

[16] P. Ayton and E. Pascoe. Bias in human judgment under uncertainty? *The knowledge engineering review*, 10(1):21–41, 1995.

[17] B M Ayyub. *Elicitation of experts opinions for uncertainty and risks.* CRC Press, 2001.

[18] J Brown, W Edwards, A Murphy and R L Winkler B Abramson. Hailfinder: A bayesian system for forecasting severe weather. *International journal of forecasting*, 12:57–71, 1996.

[19] O Bangso and P H Wuillemin. Object oriented bayesian networks: A framework for topdown specification of large bayesian networks and repetitive structures. Technical report, University of Aalborg, 2000.

[20] J M Bernardo. Bayes statistics. *Encyclopedia of Life Support Systems (EOLSS) Probability and Statistics, (R Viertl, edition) Oxford, UK*, 2003.

[21] T. Blunden. *Operational Risk: Regulation, Analysis, and Management*, chapter 11. Scorecards approaches. Prentice Hall, 2003.

[22] F. Bolger. Cognitive expertise research and knowledge engineering. *The knowledge engineering review*, 10(1):3–19, 1995.

[23] T. Bove. *Development and Validation of Human Error Management Taxonomy in Air Traffic Control.* PhD thesis, University of Roskilde, 2002.

[24] B. G. Buchanan and E. H. Shortliffe. *Rule Based expert systems: The MYCIN experiments of the stanford heuristic programming project.* Addison-Wesley, 1984.

[25] L. Cadbury. The financial aspects of corporate governance. Technical report, Cadbury Report, 1992.

[26] Cannon-Bower and Blickensderfer. Methods for eliciting declarative knowledge (tables). Technical report, NATO Defence group, 1993.

[27] R. Carnap. Logical foundations of probability. *The Internet encyclopedia of philosophy.*, 1950.

[28] S. M. Casey. *Set Phasers on Stun and Other true tales of design, Technology, and human error.* Aegean publishing, 1998.

[29] H. Chan and A. Darwiche. When do numbers really matter? In *Journal of Artificial Intelligence Research 17*, pages 265–287, 2002.

[30] P. Checkland. Towards a system-based methodology for real-world problem solving. *Journal of System Engineering*, 3(2):87–116, 1981.

[31] P. Checkland and J. Poulter. *Learning for action.* John Wiley and Sons ltd, 2006.

[32] P. Checkland and J. Scholes. *Soft Systems. Methodology in Action.* John Wiley and Sons ltd, 1990.

[33] D. N. Chorafas. *Operational risk control with Basel II. Basic principles and capital requirements.* Elsevier, 2004.

[34] M. G. Christel and K. C. Kang. Issues in requirement elicitation. Technical report, Carnegie Mellon University, 1992.

[35] P. F. Christoffersen. *Elements of Financial Risk Management.* Academic Press, 2003.

[36] P. Congdon. *Applied Bayesian Modelling*. Wiley. 2003.

[37] SERENE Consortium. *SERENE (Safety and Risk Evaluation using Bayesian Nets): Method Manual*, 1999.

[38] G. F. Cooper. The computational complexity of the probabilistic inference using bayesian belief networks. In *Artificial Intelligence an International Journal*, volume 41-42, 1989.

[39] P. Costa and K. Laskey. Pr-owl: A framework for probabilistic ontologies. In *Proceedings of the International Conference on Formal Ontologies in Information Systems, FOIS'06*, 2006.

[40] Veerle M.H. Coupé and Linda C. van der Gaag. Practicable sensitivity analysis of bayesian belief networks. In *Prague Stochastics '98 - Proceedings of the Joint Session of the 6th Prague Symposium of Asymptotic Statistics and the 13th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Union of Czech Mathematicians and Physicists*, pages 81–86, 1998.

[41] Veerle M.H. Coupé and Linda C. van der Gaag. Properties of sensitivity analysis of bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, 36(4):323–356, 2002.

[42] F. Cozman and E. Krotkov. Truncated gaussians as tolerance sets. technical report cmu-ri-tr-94-35. Technical report, Carnegie Mellon University, Pittsburgh,, 1994.

[43] M. G. Cruz. *Modelling, measuring and hedging operational risk*. John Wiley and Sons Ltd., 2002.

[44] L. Cullen. The ladbrokegrove rail enquiry. Technical report, Health and Safety Commission, 2001.

[45] C. V. Currie. *Operational Risk Modelling and Analysis - Theory and Practice.*, chapter 4. Basel II and Operational Risk - An Overview. Risk Books, 2004.

[46] Jenkinson D. The elicitation of probabilities - a review of the statistical literature. BEEP working paper, 2005. http://www.shef.ac.uk/content/1/c6/03/09/33/review3.pdf.

[47] M. Wellman D. Heckerman, A. Mamdani. Real-world applications of bayesian networks. *Comm ACM*, 38(3):25–26, 1995.

[48] da Costa P C G and Laskey K B. Multi-entity bayesian networks without multi-tears. Department of Systems Engineering and Operations Research, 2005.

[49] G. D'Agostini. Overcoming priors anxiety. *Special Issue of Revista de la Real Academia de Ciencias.*, 93(3), 1999.

[50] B. G. Dale. *Total Quality Management.* Blackwell Business, 1996.

[51] Heckerman DE and Nathwani BN. Toward normative expert systems: Part ii. probability-based representations for efficient knowledge acquisition and inference. *Methods of information in medicine*, 31(2):106–16, 1992.

[52] Bruno De Finetti. *The theory of probability.* John Wiley and sons ltd. New York, 1974.

[53] P. G. Dempsey. *Handbook of human factors and ergonomics*, chapter 41. Accident and Incident Investigation. Gavriel Salvendy, 2006.

[54] Derived by the Committee on Corporate Governance from the Committee's Final Report from the Cadbury and Greenbury Reports. *The Combined Code. Principles of Good Governance and Code of Best Practice*, 2001.

[55] J. Diez. *Razonamiento Aproximado.* http://www.ia.uned.es/fjdiez/libros/razaprox.pdf, 1989.

[56] Ph.D Douglas A. Wiegmann, Ph.D., Scott A. Shappell, Ph.D, Albert Boquet. Human error and general aviation accidents: A comprehensive, fine-grained analysis using hfacs. Technical report, Federal Aviation Administration Atlantic City International Airport, NJ, 2005.

[57] M. J. Druzdzel, Agnieszka Onisko, Daniel Schwartz, John N. Dowling and Hanna Wasyluk. Knowledge engineering for very large decision-analytic medical models. In *In Proceedings of the 1999 Annual Symposium of the American Medical Informatics Association (AMIA-1999)*, page 1049, 1999.

[58] C. G. Drury. *Handbook of human factors and ergonomics*, chapter 42. Human factors and ergonomics audits. Gavriel Salvendy, 2006.

[59] M. Druzdzel and H. Simon. Causality in bayesian belief networks. In *Proceedings of Ninth Conference on Uncertainty in Artificial Intelligence, Washington, DC*, pages 3–11. Morgan Kaufmann, 1993.

[60] M. J. Druzdzel. Explanation in probabilistic systems: Is it feasible? will it work? In *Intelligent Information Systems.*, pages 1–13. 5th Proceedings of the Workshop held in Deblia, Poland, 1996.

[61] M. J. Druzdzel and L. van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. In *Proceedings of the 11th annual conference on uncertainty in artificial intelligence*, 1995.

[62] J. M. Gutierrez E. Castillo and A. S. Hadi. *Expert Systems and Probabilistic Network Models.* Springer-Verlag. New York, Inc., 1996.

[63] N. Balakrishnan E. Castillo, A. S. Hadi and J. M. Sarabia. *Extreme Value and Related Models with Applications in Engineering and Science.* Wilet, 2005.

[64] Controlling Pilot Error. *Daryl Raymond Smith, David Frazier, L W Reithmaier, and James C Miller.* McGraw-Hill Professional, 2001.

[65] M. Druzdzel et al. Genie and smile. http://genie.sis.pitt.edu/, 2001.

[66] Bonnefon J F and Villejoubert G. Communicating likelihood and managing face: Can we say it is probable when we know it to be certain? In L. Barsalou B. G. Bara and M. Buciarelli, editors, *Proceedings of the 27th Annual Conference of the Cognitive Science Society (CogSci 2005)*, pages 316–321, 2005.

[67] Stephen Fay. *The Collapse of Barings.* W W Norton, 1997.

[68] D Fensel. Ontologies: Silver bullet for knowledge management and electronic commerce, 2001.

[69] N. Fenton. On-line tutorial on probability theory and bayesian belief bayesian networks. Queen Mary part of the University of London.

[70] N. Fenton. Incorporating expert judgement in operational risk quantification. In *Critical systems conference*. 2001.

[71] N. Fenton, P. Krause, and M. Neil. Software measurement: Uncertainty and causal modeling. *IEEE Softw.*, 19(4):116–122, 2002.

[72] B. Fischoff. *In Judgement under uncertainty heuristics and biases*, chapter Debiasing. Cambridge University Press, 1982.

[73] N. Fenton, W. Marsh, M. Neil, P. Cates, S. Forey and T. Tailor. Making resource decisions for software projects. In IEEE Computer Society, editor, *26th International Conference on Software Engineering*, pages 397–406, 2004.

[74] D. G. Fryback. Baye's theorem and conditional non-independence of data in medical diagnosis. *Computers and Biomedical Research*, 11:423–434, 1978.

[75] Gingerenzer G and Hoffrage U. How to improve bayesian reasoning without instruction: Frequency formats. *Psychological review*, 102:684–704, 1995.

[76] Wright G and Ayton P. Eliciting and modeling expert knowledge. *Decision Support Systems*, 3, 1987.

[77] Wright G and Ayton P. Decision time, subjective probability and task difficulty. *Memory and Cognition*, 16:176–85, 1988.

[78] S. Gadd and A. M. Collins. Safety culture: A review of the literature. Technical report, Health and Safety laboratory, 2002.

[79] Kadane J B Garthwaite P H and O'Hagan A. Statistical methods for eliciting probability distributions. *American Statistical Association*, 100:680–701, 2005.

[80] Dan Geiger and David Heckerman. Separable and transitive graphoids. In *UAI '90: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 65–76. Elsevier Science Inc., 1991.

[81] J. Giarratano and G. Riley. *Expert Systems: Principles and Programming*. PWS Publishing Co., third edition edition, 2001.

[82] G. Gigerenzer. *Subjective Probability*, chapter 7. Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). J.Wiley and Sons Ltd., 1994.

[83] J. Gill. *Bayesian Methods. A social and behavioural sciences approach.* Chapman and Hall/Crc., 2002.

[84] P. Giudici and A. Bilotta. Modelling operational losses: A bayesian approach. *Quality and Reliability Engineering International*, 20:407–417, 2004.

[85] T R Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.

[86] T R Gruber. Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43:907–928, 1995.

[87] Murphy A H and Winkler R L. Probability forecasting in meteorology. *American Statistical Assoc*, 1984.

[88] D. Dash H. Wang and M.J. Druzdzel. A method for evaluating elicitation schemes for probabilities. In *Artificial Intelligence Research Society Conference. Proceedings of the 13th International Florida*, 2000.

[89] M. Haas and T. Kaiser. Prerequisites for operational risk decision and management. *Operational risk*, 6, 2005.

[90] J. Haggerty. Banks struggle with operational risk. *Financial Review*, August 2001.

[91] A. Hajek. Interpretations of probability. *The Stanford Encyclopedia of Philosophy.*, 2003.

[92] R M Hamm. Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior and Human Decision Processes*, 48:193–223, 1991.

[93] R M Hamm. Characterizing individual strategies illuminates nonoptimal behavior. *Psycoloquy*, 9(49), 1998.

[94] S. Hawking. *A Brief History of Time. From the Big Bang to the Black Holes*. Bantam Books. 1988.

[95] Health and Safety Executive (HSE). Reducing risks, protecting people. *Decision making process*, 2001.

[96] Horvitz, E. J. Heckerman, D. E. and B. N. Nathwani. Toward normative expert systems: Part i the pathfinder. *Methods of Information in Medecine*, 31:90–105, 1992.

[97] E. M. Helsper and L. C. van der Gaag. Building bayesian networks through ontologies. In *Proceedings of the 15th European Conference on Artificial Intelligence*, pages 680–684. F. van Harmelen, 2002.

[98] Eveline M. Helsper and Linda C. Van Der Gaag. Ontologies for probabilistic networks: a case study in the oesophageal-cancer domain. *Knowl. Eng. Rev.*, 22(1):67–86, 2007.

[99] Max Henrion, Malcolm Pradhan, Brendan Del Favero, Kurt Huang, Gregory Provan, and Paul O'Rorke. Why is diagnosis using belief networks insensitive to imprecision in probabilities/? In *Uncertainty in Artificial Intelligence*, pages 307–314, 1996.

[100] D. G. Hoffman. *Managing Operational Risk*. John Wiley and Sons, Inc., 2002.

[101] Ulrich Hoffrage and Ralph Hertwig. Hindsight bias: A price worth paying for fast and frugal memory. Sonderforschungsbereich 504 Publications 99-22, Sonderforschungsbereich 504, Universitï¿½ Mannheim, Jul 1998.

[102] E. Hollnagel. *Handbook of human factors and ergonomics*, chapter 14. Task Analysis: Why, What, and How? Gavriel Salvendy, 2006.

[103] M. Pradhan, M. Henrion, G. Provan, B. del Favero, K. Huang. The sensitivity of belief networks to imprecise probabilities: an experimental investigation. *Artificial intelligence*, 85, 1995.

[104] Thomas Lubnau II, Randy Okray, and Thomas Lubnau. *Crew Resource Management for the Fire Service*. PennWell Books, 2004.

[105] The Basel Committee in Banking Supervision. Second quantitative impact study: Loss data. Technical report, Bank for International Settlements, 2001.

[106] The Basel Committee in Banking Supervision. Sound practices for the management and supervision of operational risk. Technical report, Bank for international settlements, 2003.

[107] Druzdel M J. Qualitative verbal explanations in bayesian belief networks. *Artificial Intelligence and Simulation of Behaviour Quarterly*, 94:43–54, 1996.

[108] Pearl J. Graphical models, causality, and intervention. *Statistical Science*, 8:266–69, 1993.

[109] E. T. Jaynes. *Papers on probability, statistics and statistical physics.*, chapter Prior Probabilities, pages 116–130. Kluwer academic, 1989.

[110] E. T. Jaynes. *Probability Theory: The Logic of Science.* not published, 1994.

[111] F. V. Jensen. *An Introduction to Bayesian Networks.* UCL Press, 1996.

[112] F. V. Jensen. *Bayesian networks and decision graphs.* Springer-Verlag, 2001.

[113] M. Jones and G. Sutherland. Implementing turnbull: A boardroom briefing. Technical report, The Institute of Chartered Accountants in England and Wales, 1999.

[114] P. Jorion. *Financial risk manager. Handbook.* Wiley, second edition, 2001.

[115] Ngwenyama O K and Bryson N. Generating belief functions from qualitative preferences: an approach to eliciting expert judgmentys and deriving probability functions. *Data and knowledge engineering*, 28:145–59, 1998.

[116] J. Kadane. Contents of volume 47, 1998. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(4):738–739, 1998.

[117] J. Kadane and L. J. Wolfson. Experiences in elicitation. *The Statistician*, 47(1), 2002.

[118] A. Kalhoff and M. Haas. *Operational Risk Modelling and Analysis - Theory and Practice.*, chapter 1. Operational Risk - Management Based on the current loss data situation. Risk Books, 2004.

[119] J. L. King. *Operational Risk. Measurement and Modelling.* Wiley. 2001.

[120] T. A. Kletz. *Learning from accidents.* Gulf professional publishing, 1994.

[121] J.J. Koehler. The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19:1:1-53, 1996.

[122] Daphne Koller and Avi Pfeffer. Object-oriented bayesian networks. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence (UAI-97)*, pages 302–313, 1997.

[123] KPMG. The final frontier: managing operating risk. *The Canadian Treasurer*, 1998.

[124] PJ Krause and DA Clark. *Representing uncertain knowledge: An artificial intelligence approach.* Kluwer Academic Publishers, 1993.

[125] J Krueger. The bet on bias: a foregone conclusion? *Psycoloquy*, 9(46), 1998.

[126] M. Kynn. *Eliciting Expert Knowledge for Bayesian Logistic Regression in Species Habitat Modelling.* PhD thesis, Queensland University of Technology, Brisbane Australia, 2005.

[127] Winkler R L. The assessment of prior distributions in bayesian analysis. *American Statistical Association,*, 62:776–800, 1967.

[128] Carmen Lacave and Francisco J. Diez. A review of explanation methods for heuristic expert systems. *Knowl. Eng. Rev.*, 19(2):133–146, 2004.

[129] S. L. Lauritzen and D. J. Spiegelhalter. *Local computations with probabilities on graphical structures and their application to expert systems.* Morgan Kaufmann Publishers Inc., 1990.

[130] E. L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks.* San Francisco: Holden-Day, Inc., 1975.

[131] Wen-Chin Li and Don Harris. Pilot error and its relationship with higher organizational levels: Hfacs analysis of 523 accidents. *Aviation, Space, and Environmental Medicine*, 77(10):1056–1062, 2006.

[132] S. Lipschutz and M. Lipson. *Probability*. Schaum's outlines, second edition, 2000.

[133] Chapman Loren and Chapman Jean. Test results are what you think they are. *Psychology today*, 18-22:106–110, 1971.

[134] Agena ltd. Agenarisk software. www.agenarisk.com, 2006.

[135] B. Malcolm M. Neil and R. Shaw. Modelling an air traffic control environment using bayesian belief networks. In *21st International System Safety Conference*, Ottawa, Ontario, Canada, 2003.

[136] N. Fenton M. Neil and P. Krause. Software quality prediction using bayesian networks in software engineering with computational intelligence. In Khoshgoftaar T.M., editor, *The Kluwer International Series in Engineering and Computer Science*, 2003.

[137] N. Fenton M. Neil and L. Nielsen. Building large-scale bayesian networks. *The Knowledge Engineering Review*, 15(3):257–284, 2000.

[138] R. Shaw, S. Johnson, B. Malcolm, I. Donald, M. Neil and C. Qui Xie. Measuring and managing culturally inspired risk current issues in safety-critical systems. In *Proceedings of the Eleventh Safety-critical Systems Symposium Great Britain*. Springer, 2003.

[139] S. Forey M. Neil, N. Fenton and R. Harry. Using bayesian belief networks to predict the reliability of military vehicles. *IEE Computing and Control Engineering*, 12(1):11–20, 2001.

[140] D. J.C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press First Edition, 2003.

[141] A Marc and H Raiffa. *Judgement under uncertainty heuristics and biases*, chapter A Progress Report on the Training of Probability Assessors. Cambridge University Press, 1982.

[142] P.S. Maybeck. *Stochastic Models, Estimation and Control*, volume 1, chapter 1. Introduction. Academic Press, 1979.

[143] Tolson Messenger. Turnbull report summary. ., 1999.

[144] M. A. Meyer and J. M. Booker. *Eliciting and Analysing Experts Judgement.* The American Statistical Association and the Society for Industrial and Applied Mathematics., 2001.

[145] J. K. Mitchell, editor. *The long road to recovery: Community responses to industrial disaster.* The United Nations University, 1996.

[146] T. M. Mitchell. *Machine Learning.* McGraw Hill International Editions, 1997.

[147] M. Moscadelli. The modelling of operational risk: experience with the analysis of the data collected by the basel committee. Technical report, Banca D'Italia, 2004.

[148] M. Neil N. Fenton and J. Galan Caballero. Using ranked nodes to model qualitative judgements in bayesian nets. *IEEE Transactions on Knowledge and Data Engineering*, 19(10), 2007.

[149] M. Neil N. Fenton and A. Joseph. Predicting football results using bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, 2005.

[150] S. Nadkarni and P. P. Shenoy. A causal mapping approach to constructing bayesian networks. *Decision Support Systems*, 38:259–281, 2004.

[151] M. Neil, P. Hearty, D. Marquez, W. Marsh, P. Krause N.E. Fenton and R. Mishra. Predicting software defects in varying development lifecycles using bayesian nets. *Information and Software Technology*, 49(49):32–43, 2007.

[152] R. E. Neapolitan. *Learning Bayesian Networks.* Northeastern Illinois University. Chicago, Illinois., 2000.

[153] Fenton N Neil M and Tailor M. Using bayesian networks to model expected and unexpected operational losses. *Risk Analysis: An International Journal*, 25(4):963–972, 2005.

[154] University of Kansas. Bayesian network tools in java. http://bnj.sourceforge.net/, 2001.

[155] A. O'Hagan. Research in elicitation. Technical report. University of Sheffield, UK. Research Report: 557/05, 2005.

[156] A. O'Hagan. *Research in elicitation. In Bayesian Statistics and its Applications.* S. K. Upadhyay, U. Singh and D. K. Dey, 2006.

[157] D. O'Hare. Cognitive functions and performance shaping factors in aviation accidents and incidents. *The int. journal of Aviation psychology*, 16(2):145–156, 2006.

[158] M. Julia Flores Olav Bangsøand Finn V. Jensen. Plug and play object oriented bayesian networks. In *Selected Papers from the 10th Conference of the Spanish Association for Artificial Intelligence (CAEPIA03)*, volume 3040/2004 of *Lecture Notes in Computer Science*, pages 457–467. Springer Berlin / Heidelberg, 2004.

[159] K. Ramadurai, T. Beck, G. Scott, K. Olson and D. Spring. Operational risk management and basel ii implementation: Survey results. Technical report, Fitch Ratings, 2004.

[160] The Basel Committee on Banking Supervision. Framework for the evaluation of internal control systems. Technical report, Bank of International Settlements, 1998.

[161] The Basel Committee on Banking Supervision. Operational risk management. Technical report, Bank of International Settlements, 1998.

[162] The Basel Committee on Banking Supervision. Core principles for effective banking supervision. Technical report, Bank of International Settlements, 1999.

[163] The Basel Committee on Banking Supervision. Consultative document. operational risk. Technical report, Bank for International Settlements, 2001.

[164] The Basel Committee on Banking Supervision. Working paper on the regulatory treatment of operational risk. Technical report, Bank of International Settlements, 2001.

[165] The Basel Committee on Banking Supervision. The qualitative impact study for operational risk: Overview of individual loss data and lessons learnt. Technical report, Bank for International Settlements, 2002.

[166] The Basel Committee on Banking Supervision. Operational risk - rules language. Technical report, Bank of International Settlements, 2003.

[167] The Basel Committee on Banking Supervision. Third quantitative impact study. Technical report, Bank for International Settlements, 2003.

[168] Branch Technical Position on the Use of Expert Elicitation in the High-Level Radioactive Waste Program. Nureg-1563. Technical report, Division of Waste Management Office of Nuclear Material Safety and Safeguards. U.S. Nuclear Regulatory Commision. Washington, DC 2055-0001, 1996.

[169] Agnieszka Onisko. *Probabilistic Causal Models in Medicine. Application to diagnosys of liver disorders.* PhD thesis, Institute of Biocybernetics and Biomedical Engineering, 2003.

[170] Krause P. Learning probabilistic networks. *The knowledge engineering review*, 13:321–351, 1998.

[171] D Barbara, K B Laskey, E J Wright, G Alghamdi, S Mirza, M Revankar P C G Costa and T Shackelford. Dtb project: A behavioral model for detecting insider threats. In *International Conference on Intelligence Analysis*, 2005.

[172] J. B. Kadane P. H. Garthwaite and A. O'Hagan. Elicitation. *Journal of American statistical association*, 2004.

[173] S. Parsons and A. Saffiotti. The qualitative verification of quantitative uncertainty. In *IPMU*, pages 180–189, 1994.

[174] Gaya Carlton, Lela Holden, Patricia Hinton-Walker and Patricia W. Stone. The intersection of patient safety and nursing research. *Annual Review of Nursing Research*, 24:3–15, 2006.

[175] Anthony O'Hagan Paul H. Garthwaite. Quantifying expert opinion in the uk water industry: An experimental study. *The Royal Statistical Society: Series D (The Statistician)*, 49:455–477, 2000.

[176] J. Pearl. Reverend bayes on inference engines: A distributed hierarchical approacha computational model for causal and diagnostic reasoning in inference engines. In

In Proceedings of the National Conference on Artificial Intelligence, pages 133–136, 1983.

[177] J Pearl. Fusion, propagation, and structuring in belief networks. *Artif. Intell.*, 29(3):241–288, 1986.

[178] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference.* Morgan Kaufmann, 1988.

[179] J. Pearl. Reasoning with cause and effect. In *IJCAI '99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 1437–1449. Morgan Kaufmann Publishers Inc., 1999.

[180] J. Pearl. Bayesianism and causality, or, why i am only half-bayesiam. In D. Corfield and J. Williamson, editors, *Foundations of Bayesianism*, volume 24 of *Applied Logic Series*, pages 19–36. Kluwer Academic, 2001.

[181] J. Pearl and T. S. Verma. A theory of inferred causation. *Principles of Knowledge Representation and Reasoning*, 1991.

[182] Charles Perrow. *Normal Accidents: Living with High-risk Technologies.* Basic Books, 1984.

[183] Joseph B. Kadane, James M. Dickey, Robert L. Winkler, Wayne S. Smith, Stephen C. Peters. Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, 75(372):845–854, 1980.

[184] J. Pézier. *Operational Risk: Regulation, Analysis, and Management*, chapter 4. A constructive review of the Basel proposal on operational risk. Prentice Hall, 2003.

[185] S. J. Press. *Subjective and objective Bayesian statistics. Principles, models and applications.* Wiley, second edition, 2003.

[186] PricewaterhouseCooper's. The world of risk: part one. *PricewaterhouseCooper's global monthly on line business magazine, Executive Perspectives*, 2003.

[187] Agena Ltd. RADAR and DERA. Tracs. Technical report, Funded by Defence Evaluation Research Agency (Chertsey), Project LSF/E20262, Aug 1997 - March 1999.

[188] CIty University Radar Group, BAe, Intelectual Capital Services, Kennedy Carter, UMIST. Simp: Systems integration for major projects. EPSRC GR/N39234, 2000-2003.

[189] F. P. Ramsey. *Truth and probability. The Foundations of Mathematics and Other Essays.* R. Braithwaite, 1931.

[190] J. Reason. *Managing the Risk of Organisational Accidents.* Ashgate Publishing Ltd., 1997.

[191] J. Reason. Human error: models and management. *BMJ - British Medical Journal*, 320:768–770, 2000.

[192] J. Reason. *Managing the Risks of the Organizational Accidents.* Ashgate, 2000.

[193] S. Renooij and C. L. M. Witteman. Talking probabilities: communicating probabilistic information in words and numbers. *Journal of Approximate Reasoning*, 22(3):169–194, 1999.

[194] Staff report. Financial oversight: The securities and exchange commission (sec) and private-sector watchdogs. Technical report, Senate Committee on Governmental Affairs, 2002.

[195] The Greenbury Report. Director's remuneration . *http://www.ecgi.org/codes/country_pages/codes_uk.htm*, 1995.

[196] Peter E. Hart Richard O. Duda and David G. Stork. *Pattern Classification, 2^{nd} Edition.* Wiley Interscience, 2000.

[197] S Russell and P Norvig. *Artificial Intelligence a Modern Approach. 2nd Edition.* Prentice Hall, 2003.

[198] Mahomey S. and Laskey K. Network engineering complex belief networks. In *25th Conference on Uncertainty in AI*, pages 389–396, 1996.

[199] Mehran Sahami, Susan Dumais, David Heckerman, and Eric Horvitz. A bayesian approach to filtering junk E-mail. In *Learning for Text Categorization: Papers from the 1998 Workshop.* AAAI Technical Report WS-98-05, 1998.

[200] L. J. Savage. *Studies in Bayesian Econometrics and statistics.*, chapter Expectations of personal probabilities and expectations, pages 111–156. North-Holland/American Elsevier, 1975.

[201] S. Scandizzo. Risk mapping and key risk indicators. *Economic notes by Banca Monte dei Paschi de Siena SpA*, 34:231–256(26), 2005.

[202] Olaf Schröder, Claus Möbus, Jörg Folckers, and Heinz-Jürgen Thole. Supporting the construction of explanation models and diagnostic reasoning in probabilistic domains. In *ICLS '96: Proceedings of the 1996 international conference on Learning sciences*, pages 60–67. International Society of the Learning Sciences, 1996.

[203] K. B. Schwamb. Mental models: A survey, 1990.

[204] G. Shafer. Dempster-shafer theory. http://www.glennshafer.com/assets/downloads/article48.p

[205] S. A. Shappell and D. A. Wiegmann. Applying reason: the human factors analysis and classification system (hfacs). *Human Factors and Aerospace Safety*, 1, 2001.

[206] E. H. Shortliffe. *Computer Based Medical Consultations: MYCIN.* American Elsevier Co. Inc., 1976.

[207] R. Simons. How risky is your company? *Harvard Business Review*, pages 85–95, 1999.

[208] D. S. Sivia. *Data Analysis: A Bayesian Tutorial.* Oxford Science Publications, 1996.

[209] P. Smets. *Advances in the Dempster-Shafer Theory of Evidence*, chapter What is Dempster-Shafer's model?, pages 5–34. Wiley, 1994.

[210] D. Stone. Reaping the advantages of basel ii. *Operational risk*, 3, 2002.

[211] D. A. Wiegmann T. L. von Thanden and S. A. Shappell. Organizational factors in commercial aviation accidents. *The int. journal of Aviation psychology*, 16(3):239–261, 2006.

[212] J. Tian. Studies in causal reasoning and learning. Technical report, Department of Computer Science. University of California., 2002.

[213] H. Zhang T.L. von Thanden, D.A. Wiegmann, A.A. Mitchell, G. Sharma. Safety culture in a regional airline: Results from a commercial aviation safety survey. In 12^{th} *International Symposium on aviation psychology*, 2003.

[214] N. Turnbull. Implementing turnbull. a boardroom briefing. Technical report, Institute of Chartered Accountants., 1999.

[215] Spanish Universities: Albacete, Granada, Pais Vasco and UNED, and sponsored by the Ministry of Science and Education. Elvira system. http://leo.ugr.es/elvira/, 2001.

[216] Budescu D. V. and Karelitz T. M. Inter-personal communication of precise and imprecise subjective probabilities.

[217] L. van der Gaag, S. Renooij, C. Witteman, B. Aleman, and B. Taal. Probabilities for a probabilistic network: A case-study in oesophageal cancer. *Artificial Intelligence in Medicine*, 25(2):123–148, 2002.

[218] L. C. van der Gaag and V. M. H. Coupé. Sensitivity analysis for threshold decision making with Bayesian belief networks. *Lecture Notes in Computer Science*, 1792:37+, 2000.

[219] L.C. van der Gaag, S. Rennoij, C.L.M. Witteman, B.M.P.Aleman and B.G.Taal. How to elicit many probabilities. In Morgan Kauffman, editor, *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pages 647–654, 1999.

[220] P. Vinella. Operational risk and basel ii accords. Technical report, PVA international Inc., 2004.

[221] P. Vinella. Operational risk in terms of operational performance. Technical report, PVA international Inc., 2005.

[222] D. Vose. *Risk Analysis. A Quantitative Guide.* Wiley, 2003.

[223] M. P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44:257–303, 1990.

[224] M. P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artif. Intell.*, 44(3):257–303, 1990.

[225] D. West. *Towards a subjective knowledge elicitation methodology for the development of Expert Systems.* PhD thesis, School of Information Science Portsmouth Polytechnic., 1991.

[226] Benson PG Whitcomb KM, ́Onkal D and Curley SP. An evaluation of the reliability of probability judgments across response modes and over time. *Journal of Behavioral Decision Making*, 6(1):283–296, 1993.

[227] Robert Whittingham. *The Blame Machine: Why Human Error Causes Accidents.* Elsevier Butterworth-Heinemann, 2003.

[228] Wikipedia. Conjugate probability. http://en.wikipedia.org/wiki/Conjugateprior.

[229] G. Wright and P. Ayton. Subjective probability: What should we believe? In G. Wright and P. Ayton, editors, *Subjective probability*, pages 164–183. J. Wiley and Sons, 1994.

[230] M. Ikeda Y. Kitamura and R. Mizoguchi. *Expert systems. Vol 1*, chapter 6. A model-based expert system based on a domain ontology. C. T. Leondes. Academic Press., 2002.